1. **Паспорт Образовательной программы**

**«Машинное обучение и искусственный интеллект»**

|  |  |
| --- | --- |
| **Версия программы** | **2** |
| **Дата Версии** | **04.09.2020 г.** |

1. **Сведения о Провайдере**

|  |  |  |
| --- | --- | --- |
| 1.1 | Провайдер | ФГАОУ ВО «Уральский федеральный университет имени первого Президента России Б.Н. Ельцина» |
| 1.2 | Логотип образовательной организации | E:\9_Гранты\2020_Цифровые сертификаты\Шаблоны_заполнение программ\IMG-20201013-WA0003.jpg |
| 1.3 | Провайдер ИНН | **6660003190** |
| 1.4 | Ответственный за программу ФИО | **Обухов Олег Владимирович** |
| 1.5 | Ответственный должность | **Руководитель Центра ДПО ИРИТ-РТФ УрФУ** |
| 1.6 | Ответственный Телефон | **+7 922 133-52-01** |
| 1.7 | Ответственный Е-mail | [**o.v.obukhov@urfu.ru**](mailto:o.v.obukhov@urfu.ru) |

1. **Основные Данные**

|  |  |  |
| --- | --- | --- |
| **№** | **Название** | **Описание** |
| 2.1 | Название программы | **Машинное обучение и искусственный интеллект** |
| 2.2 | Ссылка на страницу программы | **Описание программы**  https://dpo-rtf.urfu.ru/machine-learning-ai  **Материалы программы**  https://exam1.urfu.ru/course/view.php?id=844  **Реквизиты входа:**  Страница логина - https://exam1.urfu.ru/login/  Логин - usr1607202002  Пароль - 33532579275 |
| 2.3 | Формат обучения | Онлайн |
|  | Подтверждение о наличии возможности реализации образовательной программы с применением электронного обучения и (или) дистанционных образовательных технологий с возможностью передачи данных в форме элементов цифрового следа | Имеется |
| 2.4 | Уровень сложности | Базовый |
| 2.5 | Количество академических часов | **72** |
|  | Практикоориентированный характер образовательной программы: не менее 50 % трудоёмкости учебной деятельности отведено практическим занятиям и (или) выполнению практических заданий в режиме самостоятельной работы (кол-во академических часов) | **54** |
| 2.6 | Стоимость обучения одного обучающегося по образовательной программе, а также предоставление ссылок на 3 (три) аналогичные образовательные программы иных организаций, осуществляющих обучение, для оценки объективности стоимости или обоснование уникальности представленной образовательной программы в случае отсутствия аналогичных образовательных программ на рынке образовательных услуг | **25 000 рублей**  Курсы-аналоги:  Курс №1 - 32 900 рублей  (<https://netology.ru/programs/big-data#/>)  Курс №2 - 132 996 рублей  (<https://skillfactory.ru/data-scientist>)  Курс №3 - 45 000 рублей  (<https://productstar.ru/analytics-python-course>) |
| 2.7 | Минимальное количество человек на курсе | **15** |
| 2.8 | Максимальное количество человек на курсе | **120** |
| 2.9 | Данные о количестве слушателей, ранее успешно прошедших обучение по образовательной программе | **30** |
| 2.10 | Формы аттестации | Итоговое контрольное тестирование |
|  | Указание на область реализации компетенций цифровой экономики, к которой в большей степени относится образовательная программа, в соответствии с Перечнем областей | Большие данные |

1. **Аннотация программы**

Программа предназначена для повышения квалификации граждан РФ в части технологий Искусственного интеллекта и машинного обучения, как одной из сквозных технологий цифровой экономики

В ходе изучения программы ставятся следующие задачи:

- изучение применения технологий искусственного интеллекта для анализа больших данных;

- определение критериев аналитических задач, решаемых с использованием больших данных;

- применение технологий обработки больших данных к решению прикладных задач.

В результате освоения программы слушатели изучат методы управления аналитическими проектами, базовые методы машинного обучения и статистического анализа данных.

Получат навыки использования языка Python для решения задач анализа данных и машинного обучения, научаться преобразовывать задачи бизнеса в формат аналитического проекта, извлекать и интегрировать данные из разнородных источников, структурированных и не структурированных данных.

Программа востребована и актуальна для категории слушателей, нацеленных на получение профессиональной квалификации Junior-специалист в области Data Science а также Бизнес-аналитик (в части BI), поскольку при ее разработке использовались данные выгруженные с сайта HH.RU в части требований, предъявляемых к данной профессиональной квалификации, тем самым программа соответствует актуальным требованиям, предъявляемым рынком труда в России.

Слушатели могут осваивать курс без предварительной подготовки, для освоения программы достаточно базовых знаний о математике в рамках школьной программы, а также базовых навыков работы с персональным компьютером на уровне уверенного пользователя

Программа разработана на основе требований профессиональных  
стандартов: «Специалист по большим данным». Проект Приказа Министерства труда и социальной защиты РФ «Об утверждении профессионального стандарта «Специалист по большим данным» (подготовлен Минтрудом России 28.11.2016); «Специалист по информационным системам». Утвержден приказом Минтруда России от 18.11.2014 № 896н; «Специалист по тестированию в области информационных технологий». Утвержден приказом Министерства труда и социальной защиты Российской Федерации от 11.04.2014 г. № 225н.

УТВЕРЖДАЮ

Заместитель проректора

по проектному обучению и

дополнительному профессиональному

образованию

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ В.А. Овчинникова

«\_\_\_\_»\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_2020 г.

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ

РОССИЙСКОЙ ФЕДЕРАЦИИ

Федеральное государственное автономное

образовательное учреждение высшего образования

УРАЛЬСКИЙ ФЕДЕРАЛЬНЫЙ УНИВЕРСИТЕТ

имени первого Президента России Б.Н. Ельцина

УТВЕРЖДАЮ

Заместитель проректора

по проектному обучению и

дополнительному профессиональному

образованию

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ В.А. Овчинникова

«\_\_\_\_»\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_2020 г.

Дополнительная профессиональная программа повышения квалификации

«Машинное обучение и искусственный интеллект»

72 часа

Екатеринбург

2020 г.

**ОБЩАЯ ХАРАКТЕРИСТИКА ПРОГРАММЫ**

**1.Цель программы**

Программа предназначена для повышения квалификации граждан РФ в части технологий Искусственного интеллекта и машинного обучения, как одной из сквозных технологий цифровой экономики.

.

**2.Планируемые результаты обучения:**

2.1.Знание (осведомленность в областях)

2.1.1. Методологии управления проектами в области машинного обучения

2.1.2. Знание основных методов визуализации и особенностей их применения

2.1.3. Знание условий и особенностей применения параметрических и непараметрических тестов в статистике

2.1.4. Знание условий и особенностей применения методов регрессионного анализа и корреляции

2.1.5. Знание одно и многофакторного анализа в статистике

2.1.6. Знание условий и особенностей применения методов классификации в машинном обучении

2.1.7. Знание условий и особенностей применения кластеризации в машинном обучении

2.1.8. Знание условий и особенностей применения методов сокращения размерности в машинном обучении

2.1.9. Знание условий и особенностей применения методов анализа текстов на естественном языке с помощью алгоритмов машинного обучения

2.2. Умение (способность к деятельности)

2.2.1. Использовать жизненный цикл аналитических проектов

2.2.2. Использовать различные средства и инструменты для загрузки данных из разнородных источников

2.2.3. Строить и оценивать качество статистических моделей и результатов эксперимента

2.2.4. Выбирать соответствующие задачи методы машинного обучения и правильно представлять различные виды данных и гипотез

2.2.5. Решать аналитические задачи с помощью методов интеллектуального анализа данных и машинного обучения

2.3.Навыки (использование конкретных инструментов)

2.3.1 Навык разработки и осуществления проектов в соответствии с концепцией жизненного цикла аналитически проектов

2.3.2 Навык преобразования и постановки задачи бизнеса в формат аналитического проекта

2.3.3 Навык извлечения и интеграции данных из разнородных структурированных и неструктурированных источников

2.3.4 Навык использования различных средств визуализации для решения аналитических задач

2.3.5 Навык использования языка Python для задачи анализа данных и машинного обучения

**3. Категория слушателей**

3.1. Образование: среднее профессиональное (в области ИТ, экономики), высшее образование.

3.2. Требования к квалификации: не предъявляются.

3.3. Требования к опыту профессиональной деятельности (в части программы повышения квалификации)- не предъявляются

3.4. Предварительное освоение иных дисциплин/курсов /модулей: слушатели могут осваивать курс без предварительной подготовки

**4.Учебный план программы «Новые производственные технологии»**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| № п/п | Наименование разделов  и дисциплин | Всего часов | В том числе | | |
| Лекции | Практика | Самостоятельная работа |
| **1** | **2** | **3** | **4** | **5** | **6** |
| 1 | Модуль 1. ВВЕДЕНИЕ В ЯЗЫК PYTHON | 20 | 5 | 5 | 10 |
| 2 | Модуль 2. ВВЕДЕНИЕ В СТАТИСТИЧЕСКОЕ ОБУЧЕНИЕ | 18 | 4 | 5 | 9 |
| 3 | Модуль 3. ИНТЕЛЛЕКТУАЛЬНЫЙ АНАЛИЗ ДАННЫХ И МАШИННОЕ ОБУЧЕНИЕ | 21 | 6 | 5 | 10 |
| 4 | Модуль 4. ВОПРОСЫ ИСКУСТВЕННОГО ИНТЕЛЛЕКТА | 13 | 3 | 3 | 7 |
|  | **Итоговая аттестация (зачет)** |  |  | | |
| **Итого:** | | **72** | **зачет** | | |

**5.Календарный план-график реализации образовательной программы**

|  |  |  |  |
| --- | --- | --- | --- |
| **№ п/п** | **Наименование учебных модулей** | **Трудоёмкость (час)** | **Сроки обучения** |
| **1** | Модуль 1. ВВЕДЕНИЕ В ЯЗЫК PYTHON | 20 | **1.11.2020 -4.11.2020** |
| **2** | Модуль 2. ВВЕДЕНИЕ В СТАТИСТИЧЕСКОЕ ОБУЧЕНИЕ | 18 | **4.11.2020 -9.11.2020** |
| **3** | Модуль 3. ИНТЕЛЛЕКТУАЛЬНЫЙ АНАЛИЗ ДАННЫХ И МАШИННОЕ ОБУЧЕНИЕ | 21 | **9.11.2020 -11.11.2020** |
| **4** | Модуль 4. ВОПРОСЫ ИСКУСТВЕННОГО ИНТЕЛЛЕКТА | 13 | **12.11.2020 -13.11.2020** |
|  | Итоговая аттестация |  | **15.11.2020** |
| **Всего:** | | **72** |  |

**6.Учебно-тематический план программы «Машинное обучение и искусственныйинтеллект »**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **№ п/п** | **Модуль / Тема** | **Всего, час** | **Виды учебных занятий** | | | **Формы контроля** |
| **лекции** | **практические занятия** | **самостоятельная работа** |
| 1 | МОДУЛЬ 1. «ВВЕДЕНИЕ В ЯЗЫК PYTHON» | 20 | 5 | 5 | 10 | Тест, самостоятельное задание |
| 1.1 | Тема 1. Установка и настройка фреймворка Anaconda | 4 | 1 | 1 | 2 |  |
| 1.2 | Тема 2. Знакомство с синтаксисом, методами и средами разработки | 4,5 | 1 | 1 | 2,5 |  |
| 1.3 | Тема 3. Структурное программирование | 3,5 | 1 | 1 | 1,5 | зачет |
| 1.4 | Тема 4. Библиотеки Matplotlib, Numpy | 3,5 | 1 | 1 | 1,5 |  |
| 1.5 | Тема 5. Библиотека Pandas. Подключения к источникам данных | 4,5 | 1 | 1 | 2,5 |  |
| 2 | МОДУЛЬ 2. «ВВЕДЕНИЕ В СТАТИСТИЧЕСКОЕ ОБУЧЕНИЕ» | 18 | 4 | 5 | 9 | Тест, самостоятельное задание |
| 2.1 | Тема 1. Вводные положения | 4,5 | 1 | 1 | 2,5 |  |
| 2.2 | Тема 2. Сравнение средних | 4,5 | 1 | 1 | 2,5 |  |
| 2.3 | Тема 3. Визуальный анализ данных | 4,5 | 1 | 2 | 1,5 |  |
| 2.4 | Тема 4. Корреляция и регрессия | 4,5 | 1 | 1 | 2,5 |  |
| 3. | МОДУЛЬ 3. «ИНТЕЛЛЕКТУАЛЬНЫЙ АНАЛИЗ ДАННЫХ И МАШИННОЕ ОБУЧЕНИЕ» | 21 | 6 | 5 | 10 | Тест, самостоятельное задание |
| 3.1 | Тема 1. Вводные положения | 4 | 1 | 1 | 2 | зачет |
| 3.2 | Тема 2. Алгоритмы классификации | 5 | 1 | 1 | 3 |  |
| 3.3 | Тема 3. Методы кластеризации данных и ассоциативные правила | 3,5 | 1 | 1 | 1,5 |  |
| 3.4 | Тема 4. Методы прогнозирования численных признаков | 3,5 | 1 | 1 | 1,5 | зачет |
| 3.5 | Тема 5. Факторный анализ и сокращение размерности | 5 | 2 | 1 | 2 |  |
| 4. | МОДУЛЬ 4. «ВОПРОСЫ ИСКУСТВЕННОГО ИНТЕЛЛЕКТА» | 13 | 3 | 3 | 7 |  |
| 4.1 | Тема 1. Введение в искусственный интеллект | 4 | 1 | 1 | 2 |  |
| 4.2 | Тема 2. Обработка естественного языка | 5 | 1 | 1 | 3 |  |
| 4.3 | Тема 3. Представление знаний | 4 | 1 | 1 | 2 |  |
|  | ИТОГО часов | 72 | 18 | 18 | 36 |  |

**7. Учебная (рабочая) программа повышения квалификации «Машинное обучение и искусственный интеллект»**

**Модуль 1. ВВЕДЕНИЕ В ЯЗЫК PYTHON ( 20 час.)**

**Тема 1.1 Установка и настройка фреймворка Anaconda** **(4 час)**

Установка фреймворка Anaconda. Настройка рабочего окружения. Знакомство со средой программирования Jupyther Notebook. Преимущества и особенности Python.

**Тема 1.2 Знакомство с синтаксисом, методами и средами разработки (4,5 час)**

Синтаксис, типы данных, преобразование типов, математические операции, ветвление, ввод/вывод в Python.

**Тема 1.3 Структурное программирование ( 3,5 час)**

Списки и работа с ними. Циклы. Генераторы списков и кортежей. Работа с файлами в Python.. Функции. Словари.

**Тема 1.4 Библиотеки Matplotlib, Numpy ( 3,5 час)**

Массивы и основные операции с ними в библиотеке Numpy. Основы визуализации с помощью библиотеки Matplotlib.

**Тема 1.5 Библиотека Pandas. Подключения к источникам данных (4,5 час)**

Объекты и основные операции с ними в библиотеке Pandas. Принципы подключения самых распространённых источников данных.

**Модуль 2. ВВЕДЕНИЕ В СТАТИСТИЧЕСКОЕ ОБУЧЕНИЕ ( 18 час.)**

**Тема 2.1 Вводные положения ( 4,5 час)**

Генеральная совокупность и выборка Типы переменных. Количественные и номинативные переменные. Меры изменчивости. Нормальное распределение. Доверительные интервалы для среднего. Идея статистического вывода, p-уровень значимости.

**Тема 2.2 Сравнение средних (4,5 час)**

Сравнение двух средних, t-критерий. Проверка распределения на нормальность. Однофакторный дисперсионный анализ Множественные сравнения в ANOVAМногофакторный ANOVA.

**Тема 2.3 Визуальный анализ данных (4,5 час)**

Квартет Энксомба. Типы визуальных диаграмм и условия их применения. Правила визуализации данных. Типичные ошибки и манипуляции с помощью визуализации.

**Тема 2.4 Корреляция и регрессия (4,5 час)**

Условия применения коэффициента корреляции. Регрессия с одной независимой переменной. Гипотеза о значимости взаимосвязи и коэффициент детерминации. Условия применения линейной регрессии с одним предиктором. Применение регрессионного анализа и интерпретация результатов. Задача предсказания значений зависимой переменной. Регрессионный анализ с несколькими независимыми переменными. Выбор наилучшей модели.

**Модуль 3. ИНТЕЛЛЕКТУАЛЬНЫЙ АНАЛИЗ ДАННЫХ И МАШИННОЕ ОБУЧЕНИЕ ( 21 час.)**

**Тема 3.1 Вводные положения ( 4 час)**

Понятие интеллектуального анализа данных (DM). Требования, предъявляемые к новым знаниям. Задачи Data Mining. Стадии Data Mining. Обзор типовых задача Data Mining: классификация, кластеризация, поиск ассоциативных правил, регрессионные модели и т. д. Сферы применения технологии интеллектуального анализа данных. Методика обнаружения нового знания в хранилищах и базах данных (KDD).

**Тема 3.2 Алгоритмы классификации ( 5 час)**

Постановка задачи классификации и представление результатов. Классификационные правила: деревья решений; методы построения деревьев решений, ансамбли классификационных алгоритмов, бэггинг и бустинг, случайный лес. логистическая регрессия, наивный байесов классификатор, машина опорных векторов (SVM), метрики качества классификации, проблема переобучения.

**Тема 3.3 Методы кластеризации данных и ассоциативные правила ( 3,5 час)**

Основные понятия ассоциативных правил, алгоритм Apriori. постановка задачи кластеризации данных, процесс кластеризации данных, представление результатов кластеризации. Агломеративные и дивизимные алгоритмы, алгоритм ближайшего соседа, представление результатов иерархической кластеризации. Итеративные алгоритмы, алгоритм k-means, метрика качества алгоритмов кластеризации, методы определения вероятного числа кластеров, методы визуализации кластерной структуры

**Тема 3.4 Методы прогнозирования численных признаков ( 3,5 час)**

Регрессионные деревья. Продвинутые регрессионные модели. Прогнозирование временных рядов методами машинного обучения.

**Тема 3.5 Факторный анализ и сокращение размерности ( 5 час)**

Постановка задачи, подготовка к факторному анализу, выделение первичных факторов, метод главных компонент, алгоритм NIPALS, аналитическое вращение факторов, ортогональное и косоугольное вращение, критерии вращения. Методы отбора переменных на основе информационных критериев.

**Модуль 4. ВОПРОСЫ ИСКУСТВЕННОГО ИНТЕЛЛЕКТА ( 13 час.)**

**Тема 4.1 Введение в искусственный интеллект ( 4 час)**

Философские основы искусственного интеллекта. Автономный интеллект. Интеллектуальные агенты. История искусственного интеллекта. Философские проблемы искусственного интеллекта. Технологическая сингулярность. Василиск Роко.

**Тема 4.2 Обработка естественного языка ( 5 час)**

Формальный и статистический подходы к обработке естественного языка. Статистические методы: языковая модель, скрытая марковская модель, алгоритм Витерби, основы лингвистики (токенизация, морфология, выделение частей речи). Парсинг (анализ интересов и зависимостей).

**Тема 4.3 Представление знаний ( 4 час)**

Семантические сети, элементы глубокого обучения для систем основанных на знаниях, онтологии и графы знаний.

**Описание практико-ориентированных заданий и кейсов**

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Номер темы/модуля** | **Наименование практического занятия** | **Описание** |
| 1.1 | Установка и настройка фреймворка Anaconda | Установка и настройка среды Anaconda | Установка и настройка среды Anaconda |
| 1.2 | Знакомство с синтаксисом, методами и средами разработки | Решение арифметических задач. Обработка ввода пользователя | Знакомство с типами данных, интерфейсом Jupyter Notebook |
| 1.3 | Структурное программирование | Реализация «петли событий». Чтение и разбор текстового набора данных из файла. Создание своего модуля и пакета из модулей участников. | Реализация циклов и функций в языке Python. Считывание csv-файла |
| 1.4 | Библиотеки Matplotlib, Numpy | Решение задач на использование библиотек Matplotlib и Numpy. | Разбор типовых методов визуализации в Matplotlib. Работа с Numpy-массивами |
| 1.5 | Библиотека Pandas. Подключения к источникам данных | Основы работы с Pandas | Структуры данных Pandas. Подключение источников данных. Группировка и объединение данных в Pandas |
| 2.1 | Вводные положения | Планирование статистического эксперимента | Теория и практика дизайна стат.исследования |
| 2.1 | Вводные положения | Анализ мощности выборки | Анализ мощности в SciPy. Способы построения выборки |
| 2.1 | Вводные положения | Типичные проблемы статистических экспериментов. Поправка Бонферрони | Тесты множественных сравнений в SciPy |
| 2.2 | Сравнение средних | Однофакторный дисперсионный анализ | Однофакторный дисперсионный анализ в Python |
| 2.2 | Сравнение средних | Множественные сравнения в ANOVA | Множественные сравнения в ANOVA c помощью Python |
| 2.2 | Сравнение средних | Многофакторный ANOVA | Многофакторный ANOVA Python |
| 2.3 | Визуальный анализ данных | Визуализация данных | Визуализация данных средствами Pandas, Plotly, Seaborn |
| 2.4 | Корреляция и регрессия | Регрессия с одной независимой переменной | Регрессия с одной независимой переменной средствами Scikit-learn |
| 2.4 | Корреляция и регрессия | Регрессионный анализ с несколькими независимыми переменными | Регрессионный анализ с несколькими независимыми переменными средствами Scikit-learn |
| 2.4 | Корреляция и регрессия | Выбор наилучшей регрессионной модели | Выбор наилучшей регрессионной модели средствами Scikit-learn |
| 2.4 | Корреляция и регрессия | Регуляризация коэффициентов регрессии и отбор информативных признаков. | Регуляризация коэффициентов регрессии и отбор информативных признаков средствами Scikit-learn |
| 3.1 | Вводные положения | Жизненный цикл аналитики по методологии CRISP-DM | Разработка жизненного цикла аналитического проекта |
| 3.1 | Вводные положения | Очистка и предобработка данных | Базовая очистка и предобработка данных средствами Pandas |
| 3.2 | Алгоритмы классификации | Классификация с помощью деревьев решений | Реализация деревьев решений средствами Scilit-learn |
| 3.2 | Алгоритмы классификации | Классификация с помощью метода naïve bayes | Реализация naïve bayes средствами Scikit-learn |
| 3.3 | Методы кластеризации данных и ассоциативные правила | Кластеризация с помощью алгоритма k-means | Реализация k-means средствами Scikit-learn |
| 3.4 | Методы прогнозирования численных признаков | Прогнозирование временных рядов | Прогнозирование временных рядов средствами Scikit-learn |
| 3.5 | Факторный анализ и сокращение размерности | Факторный анализ | Реализация t-SNE и PCA средствами Scikit-learn |
| 4.2 | Обработка естественного языка | Обработка естественного языка: модель мешка слов для классификации | Построение Bag-of-words с помощью Scikit-learn. Задача сентимент-анализа |
| 4.2 | Обработка естественного языка | Обработка естественного языка: тематическое моделирование | Построение тематически моделей с помощью Gensim. |
| 4.2 | Обработка естественного языка | Обработка естественного языка: анализ эмоциональной окраски | Классификация эмоциональной окраски отзывов банков с помощью Scikit-learn. |
| 4.2 | Обработка естественного языка | Обработка естественного языка: скрытые марковские модели и поля | Скрытые марковские модели и поля с помощью Scikit-learn. |
| 4.3 | Представление знаний | Построение графа знаний | Построение графов средствами Gephi |

**8. Оценочные материалы по образовательной программе**

**8.1. Вопросы тестирования по модулям**

|  |  |  |  |
| --- | --- | --- | --- |
| **№ модуля** | **Вопросы входного тестирования** | **Вопросы промежуточного тестирования** | **Вопросы итогового тестирования** |
| Введение | Оцените свои знания и навыки, используя шкалу от 1 до 10 (1 - никогда не слышал об этом, 10 – считаю себя экспертом в этой области)  1) Классическая статистика и статистическое обучение (центральная тенденция (и другая дескриптивная статистика), ANOVA, параметрические и непараметрические методы корреляции, простые регрессионные модели и т. Д.) - от 1 до 10  2) Визуализация и подготовка данных (базовые графики, масштабирование переменных, обнаружение и борьба с выбросами, заполнение отсутствующих данных, интерполяция отсутствующих числовых данных, транасформация категориальных данных) - от 1 до 10  3) Машинное обучение (supervised, unsupervised, semi-supervised learning) - от 1 до 10  4) Дополнительные темы в области машинного обучения (интеллектуальный анализ текста, распознавание изображений, нейронные сети, deep и reinforced learning) - от 1 до 10  5) Социальные графы (метрики графа, обнаружение сообщества, поток информации в графах) - от 1 до 10  6) Знание Python (пакетов / библиотек - pandas, sklearn и других) - от 1 до 10 |  |  |
| Модуль **1** |  | Примеры тестовых заданий   1. Какая команда позволяет изменить тип данных в колонке Pandas датафрейма 2. astype() \*\* 3. set() 4. dtype() 5. shape() 6. Какая команда Pandas позволяет узнать структуру DataFrame или Series? 7. astype() 8. set() 9. dtype() 10. shape()\*\* 11. Какая команда Pandas удаляет строки или столбцы содержащие пропуски данных? 12. dropna()\*\* 13. fillna() 14. replace()   split()   1. Зачем нужна визуализация density plot в Pandas? 2. Для того чтобы подчеркнуть форму распределения данных \*\* 3. Для того чтобы не акцентировать внимание на форме распределения данных 4. Используется, когда данные сильно смещены вправо. 5. Используется, когда данные сильно смещены влево. 6. Изучите график:   skewed1.gif  Как бы вы посоветовали улучшить данный график (используя указанные ниже инструменты) для того, чтобы найти закономерности, которые вы бы иначе могли пропустить?   1. Использовать log график (Purchase size) \*\* 2. Использовать density plot 3. Использовать гистограмы 4. Использовать график типа “ящик с усами” (box-whisker) |  |
| Модуль 2 |  | Примеры тестовых заданий   1. Сопоставьте тезисы: 2. Нулевая гипотеза (H0) верна, и вы не смогли принять нулевую гипотезу 3. Нулевая гипотеза (H0) неверна, и вы не смогли отклонить нулевую гипотезу 4. – Ошибка первого типа 5. – Ошибка второго типа 6. Рассмотрим ситуацию:   Предположим, каждый, кто посещает наш сайт, либо получает одно из двух рекламных предложений, либо вообще не получает никакой рекламы.  Мы хотим увидеть, имеет ли значение рекламное предложение.  Какой статистический метод вы бы порекомендовали для этого анализа?   1. P-test 2. T- test 3. Z-test 4. ANOVA \*\* 5. Каково предположение в t-тесте Стьюдента? 6. Выборки нормально распределены\*\* 7. Выборки не нормально распределены 8. Размер выборок мал 9. Не делает никаких  |  | | --- | | 1. Вы анализируете две нормально распределенные выборки, и ваша нулевая гипотеза состоит в том, что среднее m1 первой выборки равно среднему m2 второй. Вы наблюдали значение «p» равное 4,33e-05.   Каким будет ваше решение относительно нулевой гипотезой? |  1. p-значение мало, отклоните нулевую гипотезу и примите альтернативную гипотезу \*\* 2. p-значение мало, принимаем нулевую гипотезу и отклоняем альтернативную гипотезу 3. p-Value может не предоставлять достаточной информации и учитывать другую статистику теста 4. Вы должны преобразовать данные с помощью log () и повторить анализ. |  |
| Модуль 3 |  | Примеры тестовых заданий   1. Какая часть жизненного цикла аналитического проекта занимает большую часть времени?    1. Исследование предметной области    2. Подготовка данных \*\*    3. Обзор результатов    4. Операционализация 2. Какая часть жизненного цикла аналитического проекта позволяет сформировать начальные гипотезы?    1. Исследование предметной области \*\*    2. Подготовка данных    3. Обзор результатов    4. Операционализация 3. Для чего обычно используется визуализация, на этапе подготовки данных? 4. Моделирование данных 5. Выбор переменных 6. Очистка данных 7. Исследование 8. Какова польза от запуска пилотного проекта на заключительном этапе (операционализацияя) аналитического проекта? 9. Знание о ограничениях производительности используемых методов 10. Знание о том как дообучать модель на этапе эксплуатации 11. Ограничение рисков 12. Все вышеперечисленное\*\* 13. Что из нижеперечисленного **не является** одной из потребностей спонсора проекта, которую необходимо учесть в окончательной презентации? 14. ROI (Return on Investment) – возврат инвестиций на проект 15. Влияние на бизнес 16. Риски продолжения проекта 17. Влияние на другие связанные системы 18. Вы хотите сгруппировать элементы в наборе данных по сходству и присвоить каждой группе метки. Какой аналитический метод лучше всего использовать для этой задачи? 19. K-means кластеризация\*\* 20. Naïve Bayesian Classifier 21. Logistic Regression 22. ANOVA 23. Одна из переменных в модели линейной регрессии, которую вы строите, - это «доход», который имеет значения от 1000 до 10000000.   Что является «лучшей практикой» при построении вашей модели?   1. Использовать преобразование log(доход) вместо необработанных значений, чтобы обеспечить нормальное распределение данных. \*\* 2. Использовать преобразование exp(доход) вместо необработанных значений, чтобы обеспечить нормальное распределение данных. 3. Использовать значения дохода только в диапазоне от 100000 до 500000. 4. Данные не подходят для линейной регрессии и рекомендуктся использовать логистическую регрессию. 5. Как логистическая регрессия используется в качестве классификатора? 6. Назначает метки классов на основе предварительно установленного порога для перенастроенной вероятности класса \*\* 7. Используя значение коэффициента b0 8. С обращением логит-выражения 9. Почему наивный байесовский классификатор считается вычислительно эффективным для задач большой размерности? 10. Обрабатывает категориальные переменные с большим количеством уровней \*\* 11. Использует log вероятностей вместо сырых вероятностей для вычислений 12. Алгоритмы естественно подходят для архитектур с массовой параллельной обработкой. 13. Использует функцию сглаживания Лапласа для точного вычисления вероятностей |  |
| Модуль 4 |  | Примеры тестовых заданий   1. Какие задачи решаются с помощью анализа текста? (Выбрать все, что подходит 2. Парсинг \*\* 3. Информационный поиск \*\* 4. Text Mining \*\* 5. Настройка кластеров 6. В чем заключается основная проблема представления корпуса в анализе текста? 7. Метрики корпуса динамические \*\* 8. Расчет TF-IDF 9. Низкая размерность 10. Расчет обратного индекса 11. Какая польза от TF-IDF? 12. Он обеспечивает меру, которая оценивает присутствие необычных терминов в запросе как более высокие показатели релевантности документа, чем наличие более общих терминов \*\* 13. Он обеспечивает меру, которая оценивает присутствие необычных терминов в запросе как более низкие показатели релевантности документа, чем наличие более общих терминов. 14. Он обеспечивает меру, которая оценивает присутствие общих терминов в запросе как более высокие показатели релевантности документа, чем наличие более необычных терминов.   Он позволяет определить, сколько раз термин встречается в Корпусе. |  |
| Итоговое контрольное тестирование |  |  | Примеры тестовых заданий   1. Вы хотите сгруппировать элементы в наборе данных по сходству и присвоить каждой группе метки. Какой аналитический метод лучше всего использовать для этой задачи? 2. K-means кластеризация\*\* 3. Naïve Bayesian Classifier 4. Logistic Regression 5. ANOVA 6. Посмотрите график, который отображает значение внутренней суммы квадратов для количества кластеров (k), выбранных для конкретного набора данных в решении кластеризации k-средних.   Rplot.png  Какое оптимальное количество кластеров рекомендуется использовать для этого набора данных?   1. 4\*\* 2. 6 3. 8 4. 2 5. Какие результаты дает кластерный анализ k-средних? 6. Центроиды обнаруженного кластера и назначение каждой входной точки данных кластеру \*\* 7. Правила, которые связывают каждую входную точку данных с классом и диаметром обнаруженных кластеров. 8. Суммы квадратов для каждого обнаруженного кластера и общую кластерную дисперсию 9. Ассоциация классов для каждого элемента данных и вероятности классов 10. Вы используете категориальную переменную «Регион» в своей модели линейной регрессии с 50 возможными значениями. На сколько переменных эта переменная «Регион» должна быть расширена для вашей модели? 11. 49 \*\* 12. 51 13. 2 14. Категориальные переменные нельзя использовать с линейной регрессией 15. Какой метод следует использовать для определения коэффициентов b0, b1, .., bn в вашей модели линейной регрессии:   Y = b0 + b1x1+b2x2+….+bnxn   1. Ordinary Least squares \*\* 2. Apriori 3. Ridge и Lasso \*\* 4. Динамическое программирование 5. Что означает коэффициент bcreditscore в следующей модели логистической регрессии?   default = f(creditscore,income,loanAmount,Existingdebt)  Дано bcreditscore=-0.69  exp(*bcreditScore)* = ½ и  log(*bcreditScore*)*=* -0.161   1. Для того же income, loanAmount и Existingdebt отношение шансов дефолта уменьшается вдвое на каждый пункт увеличения кредитного рейтинга \*\* 2. Для того же income, loanAmount и Existingdebt отношение шансов дефолта удваивается на каждый пункт увеличения кредитного рейтинга. 3. Для того же income, loanAmount и Existingdebt отношение шансов дефолта уменьшается на log (-0,69) для каждого увеличения кредитного рейтинга 4. Для того же income, loanAmount и Existingdebt отношение шансов дефолта будет равно значению -0,69. 5. Перед вами стоит задача предсказать, купит ли клиент товар при посещении веб-сайта, и вероятность принятия решения о покупке. Вам будут предоставлены другие соответствующие переменные, связанные с проблемой. Какой аналитический метод вы бы порекомендовали? 6. Logistic Regression \*\* 7. Linear Regression 8. ANOVA 9. Association Rules 10. Какая из следующих формул соответствует теореме Байеса? 11. P(B|A) = P(A|B)\*P(B)/P(A) \*\* 12. P(B|A) = P(A|B)\*P(A)/P(B) 13. P(B | A)\*P(B) = P(A | B)\*P(B) 14. P(A| B)\*P(B) = P(A | B)\*P(B) 15. Во входных данных присутствуют нелинейные данные и неоднородности, которые, как вы подозреваете, могут повлиять на выходные данные вашей модели. Какой аналитический метод вы предложите? 16. Decision Trees\*\* 17. ANOVA 18. K-means Clustering 19. Linear Regression 20. Какое одно из ограничений при выборе анализа временных рядов как метода прогнозирования? 21. Подходит только для краткосрочного прогнозирования \*\* 22. Невозможно удалить компонент тренда 23. Невозможно удалить компонент сезонности   ARIMA - это очень интенсивный в вычислительном отношении процесс. |

**8.2. Критерии и шкалы оценивания при освоении образовательной программы**

|  |  |
| --- | --- |
| Шкала оценивания | Критерий оценивания |
| "Зачтено" | Слушатель верно ответил на 65% и более вопросов тестовых заданий. Выполнил практические задания. В ходе выполнения практических заданий продемонстрировал умения и владения навыками применения полученных знаний и умений в ходе овладении учебным материалом |
| "Не зачтено" | Слушатель верно ответил на 64% и менее вопросов тестовых заданий. Выполнил практические задания. Не выполнил практические задания, тем самым продемонстрировав недостаточный уровень знаний и умений. |

**8.3. Примеры контрольных заданий**

**Модуль 1. ВВЕДЕНИЕ В ЯЗЫК PYTHON**

1) создать новую колонку "Sex" и заполнить ее значениями по умолчанию для каждого набора данных (moscow\_female.csv и moscow\_male.csv)

2) Объединить оба набора данных в один датафрейм (например , merge по полю "Year")

3) Выделить последнюю букву имени в именах мальчика или девочки

4) Визуализировать зависимость последней буквы в имени от пола

5) Написать функцию, которая на основе любого введеного имени, подбирает к нему обращение ("Г-жа" и "Г-дин")

**Модуль 2. ВВЕДЕНИЕ В СТАТИСТИЧЕСКОЕ ОБУЧЕНИЕ**

1) считайте данные из файла titanic.csv

2) Визуализируйте зависимость выживаемости пассажиров мужского и женского пола в зависимости от класса обслуживания

3) Выведите информацию в датафрейм о проценте выживаемости для каждого класса обслуживания пассажиров обоих полов.

**Модуль 3. ИНТЕЛЛЕКТУАЛЬНЫЙ АНАЛИЗ ДАННЫХ И МАШИННОЕ ОБУЧЕНИЕ**

1) считайте данные из файла credit\_scoring.csv

2) Обучите дерево решений с параметрами max\_depth=3 и random\_state=2020

3) Оцените точность используя метрику f1\_score

**Модуль 4. ВОПРОСЫ ИСКУСТВЕННОГО ИНТЕЛЛЕКТА**

1) считайте данные из файла sept.csv

2) Обучите тематическую модель (количество тем - 7)

3) Выведите топ-10 слов в каждой теме и дайте содержательную интерпретацию полученных тем

**8.4. Примеры тестовых заданий**

**Модуль 1. ВВЕДЕНИЕ В ЯЗЫК PYTHON**

|  |  |
| --- | --- |
| 1. Какая команда позволяет изменить тип данных в колонке Pandas датафрейма   * astype() \*\* * set() * dtype() * shape() | |
| 2. Какая команда Pandas позволяет узнать структуру DataFrame или Series?   * astype() * set() * dtype() * shape()\*\* | |
| 3. Какая команда Pandas удаляет строки или столбцы содержащие пропуски данных?   * dropna()\*\* * fillna() * replace() * split() | |

**Модуль 2. ВВЕДЕНИЕ В СТАТИСТИЧЕСКОЕ ОБУЧЕНИЕ**

|  |
| --- |
| 1. Сопоставьте тезисы:   1. Нулевая гипотеза (H0) верна, и вы не смогли принять нулевую гипотезу 2. Нулевая гипотеза (H0) неверна, и вы не смогли отклонить нулевую гипотезу 3. – Ошибка первого типа 4. – Ошибка второго типа   Правильный ответ : a –1 и b - 2 |
| 2. Рассмотрим ситуацию:  Предположим, каждый, кто посещает наш сайт, либо получает одно из двух рекламных предложений, либо вообще не получает никакой рекламы.  Мы хотим увидеть, имеет ли значение рекламное предложение.  Какой статистический метод вы бы порекомендовали для этого анализа?   * P-test * T- test * Z-test * ANOVA \*\* |
| 3. Каково предположение в t-тесте Стьюдента?   * Выборки нормально распределены\*\* * Выборки не нормально распределены * Размер выборок мал * Не делает никаких предположений относительно выборок |
| 4. Вы анализируете две нормально распределенные выборки, и ваша нулевая гипотеза состоит в том, что среднее m1 первой выборки равно среднему m2 второй. Вы наблюдали значение «p» равное 4,33e-05. Каким будет ваше решение относительно нулевой гипотезой? |

* p-значение мало, отклоните нулевую гипотезу и примите альтернативную гипотезу \*\*
* p-значение мало, принимаем нулевую гипотезу и отклоняем альтернативную гипотезу
* p-Value может не предоставлять достаточной информации и учитывать другую статистику теста
* Вы должны преобразовать данные с помощью log () и повторить анализ.

**8.5. Процедура оценки качества освоения программы**

Оценка качества освоения программы проводится в отношении соответствия результатов освоения программы заявленным целям и планируемым результатам обучения.

Оценка качества освоения программы осуществляется привлеченным лектором по программе повышения квалификации.

Контроль результатов обучения по программе включает в себя:

• промежуточную аттестацию - решение практических заданий и прохождение тестирования по итогам каждого модуля;

• итоговую аттестацию - зачет.

Итоговая аттестация слушателей проводится в форме итогового тестирования.

**9.Организационно-педагогические условия реализации программы**

**9.1. Кадровое обеспечение программы**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **№**  **п/п** | **Фамилия, имя, отчество (при наличии)** | **Место основной работы и должность, ученая степень и ученое звание (при наличии)** | **Ссылки на веб-страницы с портфолио (при наличии)** | **Фото в формате jpeg** | **Отметка о полученном согласии на обработку персональных данных** |
| 1 | Комоцкий Евгений Игоревич | Институт радиоэлектроники и информационных технологий - РТФ УрФУ, ассистент кафедры  "Аналитика больших данных и методы видеоанализа " | https://urfu.ru/ru/about/personal-pages/Personal/person/evgeny.komotsky/ |  | с обработкой персональных данных согласен |

**9.2.Учебно-методическое обеспечение и информационное сопровождение**

|  |  |
| --- | --- |
| **Учебно-методические материалы** | |
| Методы, формы и технологии | Методические разработки,  материалы курса, учебная литература |
| Интерактивные тетради с поддержкой встроенного исполнения кода на платформе Jupyter Notebook разработанные в рамках программы обучения | 1. Басараб, М.А. Интеллектуальные технологии на основе искусственных нейронных сетей [Электронный ресурс]: учебное пособие / М.А. Басараб, Н.С. Коннова. - Электрон, дан. - Москва: МГТУ им. Н.Э. Баумана, 2017. - 56 с. - Режим доступа: <https://e.lanbook.com/book/103496>. - Загл. с экрана.  2. Инженерия знаний. Модели и методы [Электронный ресурс] : учебник / Гаврилова Т. А., Кудрявцев Д. В., Муромцев Д. И., 2-е изд., стер. - :  Лань, 2018 - 324 с. - Книга из коллекции Лань - Информатика. - Доступ по логину и паролю из сети Интернет .- ISBN 978-5-8114-2128-2 .- <URL:<https://e.lanbook.com/book/107925>>.  3. Кратович, П. В. Нейронные сети и модели ARIMA для прогнозирования котировок [Текст] / П. В. Кратович // Программные продукты и системы. - 2011. - N 1. - С. 95-98: ил.: 3 рис., 3 табл. - ISSN 0236-235Х. - Библиогр.: с. 98 (3 назв.).  4. Малыгин. И. Г. Информационные технологии и искусственный интеллект  - основные двигатели четвертой индустриальной революции (Industrie 4.0) = Information Technology and Artificial Intelligence - the Main Engines of the Fourth Industrial Revolution (Industrie 4.0) / И. Г. Малыгин, В. И. Комашинский // Информационные технологии. - 2016. - Т. 22, № 12. - С. 899-904. - (Интеллектуальные системы и технологии). - ISSN 1684-6400. - Библиогр.: с. 904 (13 назв.).  5.Васильев, В. И. Искусственный интеллект в системах управления и обработки информации / В. И. Васильев // Вестник Уфимского государственного авиационного технического университета. - 2000,- №1.  - С. 133-140 .- (Научные статьи и доклады. Гуманитарные и социальные науки).- ISSN 1992-6502. - Библиогр.: с. 139-140 (26 назв.). |

|  |  |
| --- | --- |
| **Информационное сопровождение** | |
| Электронные  образовательные ресурсы | Электронные  информационные ресурсы |
| Портал СДО УрФУ - https://exam1.urfu.ru/ | Open AI. 2018. How AI Training Scales, [online] Available at: <<https://openai.com/blog/scieiice-of-ai/>> |

**9.3.Материально-технические условия реализации программы**

|  |  |
| --- | --- |
| Вид занятий | Наименование оборудования,  программного обеспечения |
| Лекции | Web-браузер, платформа Zoom (https://zoom.us/) |
| Практические занятия | Web-браузер, фреймворк Anaconda Python ver. 3.8 |

**III.Паспорт компетенций (Приложение 2)**

**ПАСПОРТ КОМПЕТЕНЦИЙ ДОПОЛНИТЕЛЬНОЙ ПРОФЕССИОНАЛЬНОЙ ОБРАЗОВАТЕЛЬНОЙ ПРОГРАММЫ ПОВЫШЕНИЯ КВАЛИФИКАЦИИ**

**ПАСПОРТ КОМПЕТЕНЦИИ 1**

**«Машинное обучение и искусственный интеллект»**

**ФГАОУ ВО «Уральский федеральный университет имени первого Президента России Б.Н. Ельцина»**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 1. | Наименование компетенции | | Организовывать работу в соответствии с концепцией жизненного цикла аналитических проектов | |
| 2. | Указание типа компетенции | общекультурная/  универсальная |  | |
| общепрофессиональная |  | |
| профессиональная | + | |
| профессионально-специализированная |  | |
| 3. | Определение, содержание и основные сущностные характеристики компетенции | | Практический опыт - Опыт разработки и осуществления проектов в соответствии с концепцией жизненного цикла аналитически проектов  Умения - Использование жизненного цикла аналитических проектов  Знания - Методология CRISP-DM | |
| 4. | Дескриптор знаний, умений и навыков по уровням | | Уровни сформированности компетенции обучающегося | Индикаторы |
|  | | Начальный уровень | Имеет фрагментарное теоретическое представление о предметной области. Не демонстрирует регулярно воспроизводимых практик, связанных с использованием навыков из знаний предметной области |
|  | | Базовый уровень | Знания  Обучающийся демонстрирует знание-знакомство, знание-копию: узнает объекты, явления и понятия, находит в них различия, проявляет знание источников получения информации, может осуществлять самостоятельно репродуктивные действия над знаниями путем самостоятельного воспроизведения и применения информации.  Умения  Обучающийся умеет корректно выполнять предписанные действия по инструкции, алгоритму в известной ситуации, самостоятельно выполняет действия по решению типовых задач, требующих выбора из числа известных методов, в предсказуемо изменяющейся ситуации  Навыки  Умеет применять навык разработки и осуществления проектов в соответствии с концепцией жизненного цикла аналитически проектов |
|  | | Продвинутый | Знания  Демонстрирует аналитические знания: уверенно воспроизводит и понимает полученные знания, относит их к той или иной классификационной группе, самостоятельно систематизирует их, устанавливает взаимосвязи между ними, продуктивно применяет в знакомых ситуациях.  Умения  Умеет самостоятельно выполнять действия (приемы, операции) по решению нестандартных задач, требующих выбора на основе комбинации известных методов, в непредсказуемо изменяющейся ситуации  Навыки  Умеет применять навык разработки и осуществления проектов в соответствии с концепцией жизненного цикла аналитически проектов в условиях недостатка первичной информации |
|  | | Профессиональный | Знания  Может самостоятельно извлекать новые знания из окружающего мира, творчески их использовать для принятия решений в новых и нестандартных ситуациях.  Умения  Умеет самостоятельно выполнять действия, связанные с решением исследовательских задач, демонстрирует творческое использование умений (технологий)  Умения  Умеет применять навык разработки и осуществления проектов в соответствии с концепцией жизненного цикла аналитически проектов в ситуациях повышенной сложности |
| 5. | Характеристика взаимосвязи данной компетенции с другими компетенциями/ необходимость владения другими компетенциями для формирования данной компетенции | | Требования не предъявляются | |
| 6. | Средства и технологии оценки | | Практические задания  Тестовые задания | |

**ПАСПОРТ КОМПЕТЕНЦИИ 2**

**«Машинное обучение и искусственный интеллект»**

**ФГАОУ ВО «Уральский федеральный университет имени первого Президента России Б.Н. Ельцина»**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 1. | Наименование компетенции | | Осуществлять процесс ETL (Extraction, Transformation, Loading) из различных видов источников данных | |
| 2. | Указание типа компетенции | общекультурная/  универсальная |  | |
| общепрофессиональная |  | |
| профессиональная | + | |
| профессионально-специализированная |  | |
| 3. | Определение, содержание и основные сущностные характеристики компетенции | | Практический опыт - Опыт извлечения и интеграции из разнородных структурированных и неструктурированных источников  Умения - Использовать различные средства и инструменты для загрузки данных из разнородных источников  Знания - Методов извлечения и агрегации данных средствами Python | |
| 4. | Дескриптор знаний, умений и навыков по уровням | | Уровни сформированности компетенции обучающегося | Индикаторы |
|  | | Начальный уровень | Имеет фрагментарное теоретическое представление о предметной области. Не демонстрирует регулярно воспроизводимых практик, связанных с использованием навыков из знаний предметной области |
|  | | Базовый уровень | Знания  Демонстрирует знание-знакомство, знание-копию: узнает объекты, явления и понятия, находит в них различия, проявляет знание источников получения информации, может осуществлять самостоятельно репродуктивные действия над знаниями путем самостоятельного воспроизведения и применения информации.  Умения  Умеет корректно выполнять предписанные действия по инструкции, алгоритму в известной ситуации, самостоятельно выполняет действия по решению типовых задач, требующих выбора из числа известных методов, в предсказуемо изменяющейся ситуации  Навыки  Умеет применять навык извлечения и агрегации данных средствами Python |
|  | | Продвинутый | Знания  Демонстрирует аналитические знания: уверенно воспроизводит и понимает полученные знания, относит их к той или иной классификационной группе, самостоятельно систематизирует их, устанавливает взаимосвязи между ними, продуктивно применяет в знакомых ситуациях.  Умения  Умеет самостоятельно выполнять действия (приемы, операции) по решению нестандартных задач, требующих выбора на основе комбинации известных методов, в непредсказуемо изменяющейся ситуации  Навыки  Умеет применять навык извлечения и агрегации данных средствами Python в случае работы с полу структурированными данными |
|  | | Профессиональный | Знания  Может самостоятельно извлекать новые знания из окружающего мира, творчески их использовать для принятия решений в новых и нестандартных ситуациях.  Умения  Умеет самостоятельно выполнять действия, связанные с решением исследовательских задач, демонстрирует творческое использование умений (технологий)  Умения  Умеет применять навык извлечения и агрегации данных средствами Python в случае работы с неструктурированными |
| 5. | Характеристика взаимосвязи данной компетенции с другими компетенциями/ необходимость владения другими компетенциями для формирования данной компетенции | | Требования не предъявляются | |
| 6. | Средства и технологии оценки | | Практические задания  Тестовые задания | |

**ПАСПОРТ КОМПЕТЕНЦИИ 3**

**«Машинное обучение и искусственный интеллект»**

**ФГАОУ ВО «Уральский федеральный университет имени первого Президента России Б.Н. Ельцина»**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 1. | Наименование компетенции | | Преобразовывать бизнес-задачи в аналитические проекты | |
| 2. | Указание типа компетенции | общекультурная/  универсальная |  | |
| общепрофессиональная |  | |
| профессиональная | + | |
| профессионально-специализированная |  | |
| 3. | Определение, содержание и основные сущностные характеристики компетенции | | Практический опыт - Опыт преобразования и постановки задачи бизнеса в формат аналитического проекта  Умения - Проведение проблемного интервью и формализация требований заказчика к аналитическому проекту  Знания - Методы работы с требованиями | |
| 4. | Дескриптор знаний, умений и навыков по уровням | | Уровни сформированности компетенции обучающегося | Индикаторы |
|  | | Начальный уровень | Имеет фрагментарное теоретическое представление о предметной области. Не демонстрирует регулярно воспроизводимых практик, связанных с использованием навыков из знаний предметной области |
|  | | Базовый уровень | Знания  Демонстрирует знание-знакомство, знание-копию: узнает объекты, явления и понятия, находит в них различия, проявляет знание источников получения информации, может осуществлять самостоятельно репродуктивные действия над знаниями путем самостоятельного воспроизведения и применения информации.  Умения  Умеет корректно выполнять предписанные действия по инструкции, алгоритму в известной ситуации, самостоятельно выполняет действия по решению типовых задач, требующих выбора из числа известных методов, в предсказуемо изменяющейся ситуации  Навыки  Умеет применять навык разработки работы с требованиями в аналитическом проекте |
|  | | Продвинутый | Знания  Демонстрирует аналитические знания: уверенно воспроизводит и понимает полученные знания, относит их к той или иной классификационной группе, самостоятельно систематизирует их, устанавливает взаимосвязи между ними, продуктивно применяет в знакомых ситуациях.  Умения  Умеет самостоятельно выполнять действия (приемы, операции) по решению нестандартных задач, требующих выбора на основе комбинации известных методов, в непредсказуемо изменяющейся ситуации  Навыки  Умеет применять навык работы с требованиями в аналитическом проекте в условиях недостатка первичной информации |
|  | | Профессиональный | Знания  Может самостоятельно извлекать новые знания из окружающего мира, творчески их использовать для принятия решений в новых и нестандартных ситуациях.  Умения  Умеет самостоятельно выполнять действия, связанные с решением исследовательских задач, демонстрирует творческое использование умений (технологий)  Умения  Умеет применять навык работы с требованиями в аналитическом проекте в ситуациях повышенной сложности |
| 5. | Характеристика взаимосвязи данной компетенции с другими компетенциями/ необходимость владения другими компетенциями для формирования данной компетенции | | Требования не предъявляются | |
| 6. | Средства и технологии оценки | | Практические задания  Тестовые задания | |

**ПАСПОРТ КОМПЕТЕНЦИИ 4**

**«Машинное обучение и искусственный интеллект»**

**ФГАОУ ВО «Уральский федеральный университет имени первого Президента России Б.Н. Ельцина»**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 1. | Наименование компетенции | | Строить статистические модели и определять их качество | |
| 2. | Указание типа компетенции | общекультурная/  универсальная |  | |
| общепрофессиональная |  | |
| профессиональная | + | |
| профессионально-специализированная |  | |
| 3. | Определение, содержание и основные сущностные характеристики компетенции | | Практический опыт - Опыт построения и оценки качества статистических моделей  Умения - Строить и оценивать качество статистических моделей и результатов эксперимента  Знания - Параметрических и непараметрических тестов в статистике, методов корреляции и регрессии | |
| 4. | Дескриптор знаний, умений и навыков по уровням | | Уровни сформированности компетенции обучающегося | Индикаторы |
|  | | Начальный уровень | Имеет фрагментарное теоретическое представление о предметной области. Не демонстрирует регулярно воспроизводимых практик, связанных с использованием навыков из знаний предметной области |
|  | | Базовый уровень | Знания  Демонстрирует знание-знакомство, знание-копию: узнает объекты, явления и понятия, находит в них различия, проявляет знание источников получения информации, может осуществлять самостоятельно репродуктивные действия над знаниями путем самостоятельного воспроизведения и применения информации.  Умения  Умеет корректно выполнять предписанные действия по инструкции, алгоритму в известной ситуации, самостоятельно выполняет действия по решению типовых задач, требующих выбора из числа известных методов, в предсказуемо изменяющейся ситуации  Навыки  Умеет применять параметрические и непараметрические тестов в статистике, методы корреляции и регрессии |
|  | | Продвинутый | Знания  Демонстрирует аналитические знания: уверенно воспроизводит и понимает полученные знания, относит их к той или иной классификационной группе, самостоятельно систематизирует их, устанавливает взаимосвязи между ними, продуктивно применяет в знакомых ситуациях.  Умения  Умеет самостоятельно выполнять действия (приемы, операции) по решению нестандартных задач, требующих выбора на основе комбинации известных методов, в непредсказуемо изменяющейся ситуации  Навыки  Умеет применять параметрические и непараметрические тестов в статистике, методы корреляции и регрессии |
|  | | Профессиональный | Знания  Может самостоятельно извлекать новые знания из окружающего мира, творчески их использовать для принятия решений в новых и нестандартных ситуациях.  Умения  Умеет самостоятельно выполнять действия, связанные с решением исследовательских задач, демонстрирует творческое использование умений (технологий)  Умения  Умеет применять параметрические и непараметрические тестов в статистике, методы корреляции и регрессии |
| 5. | Характеристика взаимосвязи данной компетенции с другими компетенциями/ необходимость владения другими компетенциями для формирования данной компетенции | | Требования не предъявляются | |
| 6. | Средства и технологии оценки | | Практические задания  Тестовые задания | |

**ПАСПОРТ КОМПЕТЕНЦИИ 5**

**«Машинное обучение и искусственный интеллект»**

**ФГАОУ ВО «Уральский федеральный университет имени первого Президента России Б.Н. Ельцина»**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 1. | Наименование компетенции | | Использовать методы интеллектуального анализа данных и машинного обучения | |
| 2. | Указание типа компетенции | общекультурная/  универсальная |  | |
| общепрофессиональная |  | |
| профессиональная | + | |
| профессионально-специализированная |  | |
| 3. | Определение, содержание и основные сущностные характеристики компетенции | | Практический опыт - Опыт использования различных алгоритмов и методов для решения аналитических задач (кластеризация, классификация и т.п.)  Умения - Решать аналитические задачи с помощью методов интеллектуального анализа данных и машинного обучения  Знания - Основных методов машинного обучения и интеллектуального анализа данных для решения аналитических задач (кластеризация, классификация и т.п.) | |
| 4. | Дескриптор знаний, умений и навыков по уровням | | Уровни сформированности компетенции обучающегося | Индикаторы |
|  | | Начальный уровень | Имеет фрагментарное теоретическое представление о предметной области. Не демонстрирует регулярно воспроизводимых практик, связанных с использованием навыков из знаний предметной области |
|  | | Базовый уровень | Знания  Демонстрирует знание-знакомство, знание-копию: узнает объекты, явления и понятия, находит в них различия, проявляет знание источников получения информации, может осуществлять самостоятельно репродуктивные действия над знаниями путем самостоятельного воспроизведения и применения информации.  Умения  Умеет корректно выполнять предписанные действия по инструкции, алгоритму в известной ситуации, самостоятельно выполняет действия по решению типовых задач, требующих выбора из числа известных методов, в предсказуемо изменяющейся ситуации  Навыки  Умеет применять навыки использования методов машинного обучения и интеллектуального анализа данных для решения аналитических задач (кластеризация, классификация и т.п.) |
|  | | Продвинутый | Знания  Демонстрирует аналитические знания: уверенно воспроизводит и понимает полученные знания, относит их к той или иной классификационной группе, самостоятельно систематизирует их, устанавливает взаимосвязи между ними, продуктивно применяет в знакомых ситуациях.  Умения  Умеет самостоятельно выполнять действия (приемы, операции) по решению нестандартных задач, требующих выбора на основе комбинации известных методов, в непредсказуемо изменяющейся ситуации  Навыки  Умеет применять навыки использования методов машинного обучения и интеллектуального анализа данных для решения аналитических задач (кластеризация, классификация и т.п.) в условиях неопределённости и нечеткой постановки задач |
|  | | Профессиональный | Знания  Может самостоятельно извлекать новые знания из окружающего мира, творчески их использовать для принятия решений в новых и нестандартных ситуациях.  Умения  Умеет самостоятельно выполнять действия, связанные с решением исследовательских задач, демонстрирует творческое использование умений (технологий)  Умения  Умеет применять навыки использования методов машинного обучения и интеллектуального анализа данных для решения аналитических задач (кластеризация, классификация и т.п.) в условиях неопределённости и нечеткой постановки задач и обосновывать применение тех или иных методов |
| 5. | Характеристика взаимосвязи данной компетенции с другими компетенциями/ необходимость владения другими компетенциями для формирования данной компетенции | | Требования не предъявляются | |
| 6. | Средства и технологии оценки | | Практические задания  Тестовые задания | |

**VI.Иная информация о качестве и востребованности образовательной программы** (результаты профессионально-общественной аккредитации образовательной программы, включение в системы рейтингования, призовые места по результатам проведения конкурсов образовательных программ и др.) - **нет**

**V.Рекомендаций к программе от работодателей**: наличие не менее двух писем и/или подтверждения на цифровой платформе Государственной системы предоставления ПЦС от работодателей о рекомендации образовательной программы для реализации в рамках Государственной системы предоставления ПЦС на формирование у трудоспособного населения компетенций цифровой экономики с указанием востребованности результатов освоения программы в сфере деятельности соответствующих компаний и готовности к рассмотрению заявок наиболее успешно освоивших образовательную программу граждан на прохождение стажировки и (или) собеседования на предмет трудоустройства путем проставления отметки в профиле программы.

- ООО «Вебтим.про»

- Свердловский областной фонд поддержки предпринимательства

**VI.Указание на возможные сценарии профессиональной траектории граждан** по итогам освоения образовательной программы (в соответствии с приложением)

**СЦЕНАРИИ ПРОФЕССИОНАЛЬНОЙ ТРАЕКТОРИИ**

|  |  |
| --- | --- |
| Цели получения персонального цифрового сертификата | |
| текущий статус | цель |
| Без статуса | развитие профессиональных качеств для получения профессиональной квалификации Junior-специалист в области Data Science |
| Без статуса | развитие профессиональных качеств для получения профессиональной квалификации Бизнес-аналитик (в части BI) |
| Без статуса | развитие профессиональных качеств для получения профессиональной квалификации Специалист по анализу данных (по отраслям) |

**VII.Дополнительная информация**

**VIII.Приложенные Скан-копии**

Утвержденная программа по шаблону УрФУ.

УТВЕРЖДАЮ

Заместитель проректора

по проектному обучению и

дополнительному профессиональному

образованию

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ В.А. Овчинникова

«\_\_\_\_»\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_2020 г.

**МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ**

**РОССИЙСКОЙ ФЕДЕРАЦИИ**

Федеральное государственное автономное

образовательное учреждение высшего образования

**УРАЛЬСКИЙ ФЕДЕРАЛЬНЫЙ УНИВЕРСИТЕТ**

имени первого Президента России Б.Н. Ельцина

УТВЕРЖДАЮ

Заместитель проректора

по проектному обучению и

дополнительному профессиональному

образованию

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ В.А. Овчинникова

«\_\_\_\_»\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_2020 г.

Дополнительная профессиональная программа

повышения квалификации

**«Машинное обучение и искусственный интеллект»**

Программа утверждена на заседании ОМО ДПО

Протокол №5 от 04.09.2020 год

Программа согласована с ОМО ДПО УрФУ

Начальник ОМО ДПО УрФУ\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_/В.Г. Вятчина

**Екатеринбург**

**2020 г.**

**1. НОРМАТИВНО-ПРАВОВЫЕ ОСНОВАНИЯ РАЗРАБОТКИ ПРОГРАММЫ**

Нормативную правовую основу разработки программы составляют:

* Федеральный закон от 29 декабря 2012 г. № 273-ФЗ «Об образовании в Российской Федерации»;
* Постановление Правительства Российской Федерации от 22 января 2013 г. № 23 «О Правилах разработки, утверждения и применения профессиональных стандартов»;
* Приказ Минтруда России от 12 апреля 2013 г. № 148н «Об утверждении уровней квалификаций в целях разработки проектов профессиональных стандартов»;
* Приказ Минобрнауки России от 1 июля 2013 г. № 499 «Об утверждении Порядка организации и осуществления образовательной деятельности по дополнительным профессиональным программам»;

**2. ЦЕЛЬ ПРОГРАММЫ -**

Целью программы является приобретение слушателями компетенций в вопросах создания систем искусственного интеллекта.

**3. ХАРАКТЕРИСТИКА ПОДГОТОВКИ ПО ПРОГРАММЕ**

3.1. Категория обучающихся, на которых рассчитана программа: разработчики IT- продуктов, специалисты в области BI и анализу данных, бизнес-аналитики.

3.2. Базовый уровень образования обучающихся: высшее образование – бакалавриат, специалитет, магистратура; среднее профессиональное образование.

3.3. Нормативный срок освоения программы *–72 часов (2з.е.).*

3.4. Режим обучения *–6-8 часов в день*

3.5*.* Форма реализации *–* онлайн

3.6. Выдаваемый документ – удостоверение о повышении квалификации.

**4. ТРЕБОВАНИЯ К РЕЗУЛЬТАТАМ ОСВОЕНИЯ ПРОГРАММЫ**

Обучающийся, освоивший программу, должен обладать **профессиональными компетенциями**, включающими в себя способность:

ПК 1) Организовывать работу в соответствии с концепцией жизненного цикла аналитических проектов;

ПК 2) Осуществлять процесс ETL (Extraction, Transformation, Loading) из различных видов источников данных;

ПК 3) Преобразовывать бизнес-задачи в аналитические проекты;

ПК 4) Строить статистические модели и определять их качество;

ПК 5) Использовать методы интеллектуального анализа данных и машинного обучения.

**5. УЧЕБНО-ТЕМАТИЧЕСКИЙ ПЛАН**

Таблица 1

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **№ п/п** | **Модуль / Тема** | **Всего, час** | **Виды учебных занятий** | | | **Формы контроля** |
| **лекции** | **практические занятия** | **самостоятельная работа** |
| 1 | МОДУЛЬ 1. «ВВЕДЕНИЕ В ЯЗЫК PYTHON» | 20 | 5 | 5 | 10 | Тест, самостоятельное задание |
| 1.1 | Тема 1. Установка и настройка фреймворка Anaconda | 4 | 1 | 1 | 2 |  |
| 1.2 | Тема 2. Знакомство с синтаксисом, методами и средами разработки | 4,5 | 1 | 1 | 2,5 |  |
| 1.3 | Тема 3. Структурное программирование | 3,5 | 1 | 1 | 1,5 | зачет |
| 1.4 | Тема 4. Библиотеки Matplotlib, Numpy | 3,5 | 1 | 1 | 1,5 |  |
| 1.5 | Тема 5. Библиотека Pandas. Подключения к источникам данных | 4,5 | 1 | 1 | 2,5 |  |
| 2 | МОДУЛЬ 2. «ВВЕДЕНИЕ В СТАТИСТИЧЕСКОЕ ОБУЧЕНИЕ» | 18 | 4 | 5 | 9 | Тест, самостоятельное задание |
| 2.1 | Тема 1. Вводные положения | 4,5 | 1 | 1 | 2,5 |  |
| 2.2 | Тема 2. Сравнение средних | 4,5 | 1 | 1 | 2,5 |  |
| 2.3 | Тема 3. Визуальный анализ данных | 4,5 | 1 | 2 | 1,5 |  |
| 2.4 | Тема 4. Корреляция и регрессия | 4,5 | 1 | 1 | 2,5 |  |
| 3. | МОДУЛЬ 3. «ИНТЕЛЛЕКТУАЛЬНЫЙ АНАЛИЗ ДАННЫХ И МАШИННОЕ ОБУЧЕНИЕ» | 21 | 6 | 5 | 10 | Тест, самостоятельное задание |
| 3.1 | Тема 1. Вводные положения | 4 | 1 | 1 | 2 | зачет |
| 3.2 | Тема 2. Алгоритмы классификации | 5 | 1 | 1 | 3 |  |
| 3.3 | Тема 3. Методы кластеризации данных и ассоциативные правила | 3,5 | 1 | 1 | 1,5 |  |
| 3.4 | Тема 4. Методы прогнозирования численных признаков | 3,5 | 1 | 1 | 1,5 | зачет |
| 3.5 | Тема 5. Факторный анализ и сокращение размерности | 5 | 2 | 1 | 2 |  |
| 4. | МОДУЛЬ 4. «ВОПРОСЫ ИСКУСТВЕННОГО ИНТЕЛЛЕКТА» | 13 | 3 | 3 | 7 |  |
| 4.1 | Тема 1. Введение в искусственный интеллект | 4 | 1 | 1 | 2 |  |
| 4.2 | Тема 2. Обработка естественного языка | 5 | 1 | 1 | 3 |  |
| 4.3 | Тема 3. Представление знаний | 4 | 1 | 1 | 2 |  |
|  | Итоговая аттестация | Зачет в форме тестирования | | | | |
|  | ИТОГО часов | 72 | 18 | 18 | 36 |  |

**6. СОДЕРЖАНИЕ ПРОГРАММЫ**

Таблица 2

|  |  |
| --- | --- |
| **Характеристика структуры обучения** | **Содержание обучения,**  **а также тематика практических занятий (семинаров, лабораторных занятий), самостоятельной работы, описание применяемых образовательных технологий и рекомендуемых методических материалов, Интернет-ресурсов** |
| **МОДУЛЬ 1. «ВВЕДЕНИЕ В ЯЗЫК PYTHON»** | |
| Тема 1. Установка и настройка фреймворка Anaconda | Установка фреймворка Anaconda. Настройка рабочего окружения. Знакомство со средой программирования Jupyther Notebook. Преимущества и особенности Python. |
| Тема 2. Знакомство с синтаксисом, методами и средами разработки | Синтаксис, типы данных, преобразование типов, математические операции, ветвление, ввод/вывод в Python. |
| Тема 3. Структурное программирование | Списки и работа с ними. Циклы. Генераторы списков и кортежей. Работа с файлами в Python. Функции. Словари. |
| Тема 4. Библиотеки Matplotlib, Numpy | Массивы и основные операции с ними в библиотеке Numpy. Основы визуализации с помощью библиотеки Matplotlib. |
| Тема 5. Библиотека Pandas. Подключения к источникам данных. | Объекты и основные операции с ними в библиотеке Pandas. Принципы подключения самых распространённых источников данных. |
| **В том числе:** | |
| Практические занятия | Установка и настройка среды Anaconda |
| Решение арифметических задач. Обработка ввода пользователя. |
| Реализация «петли событий». Чтение и разбор текстового набора данных из файла. Создание своего модуля и пакета из модулей участников. |
| Решение задач на использование библиотек Matplotlib и Numpy. |
| Основы работы с Pandas |
| Применяемые образовательные технологии | Электронный курс **https://exam1.urfu.ru/course/view.php?id=844** |
| Рекомендуемые методические материалы, Интернет-ресурсы | Материалы электронного курса https://exam1.urfu.ru/course/view.php?id=844  Электронные информационные ресурсы:   1. Open AI. 2018. How AI Training Scales, [online] Available at: <<https://openai.com/blog/scieiice-of-ai/>>   Учебно-методические материалы:  1. Басараб, М.А. Интеллектуальные технологии на основе искусственных нейронных сетей [Электронный ресурс]: учебное пособие / М.А. Басараб, Н.С. Коннова. - Электрон, дан. - Москва: МГТУ им. Н.Э. Баумана, 2017. - 56 с. - Режим доступа: <https://e.lanbook.com/book/103496>. - Загл. с экрана.  2. Инженерия знаний. Модели и методы [Электронный ресурс] : учебник /Гаврилова Т. А., Кудрявцев Д. В., Муромцев Д. И., 2-е изд., стер. - :Лань, 2018 - 324 с. - Книга из коллекции Лань - Информатика. – Доступ по логину и паролю из сети Интернет .- ISBN 978-5-8114-2128-2 .- <URL:<https://e.lanbook.com/book/107925>>.  3. Кратович, П. В. Нейронные сети и модели ARIMA для прогнозирования котировок [Текст] / П. В. Кратович // Программные продукты и системы. - 2011. - N 1. - С. 95-98: ил.: 3 рис., 3 табл. - ISSN 0236-235Х. - Библиогр.: с. 98 (3 назв.).  4. Малыгин. И. Г. Информационные технологии и искусственный интеллект- основные двигатели четвертой индустриальной революции (Industrie 4.0) = Information Technology and Artificial Intelligence - the Main Engines of the Fourth Industrial Revolution (Industrie 4.0) / И. Г. Малыгин, В. И. Комашинский // Информационные технологии. - 2016. - Т. 22, № 12. - С. 899-904. - (Интеллектуальные системы и технологии). - ISSN 1684-6400. - Библиогр.: с. 904 (13 назв.).  5.Васильев, В. И. Искусственный интеллект в системах управления и обработки информации / В. И. Васильев // Вестник Уфимского государственного авиационного технического университета. - 2000,- №1.- С. 133-140 .- (Научные статьи и доклады. Гуманитарные и социальные науки).- ISSN 1992-6502. - Библиогр.: с. 139-140 (26 назв.). |
| **Модуль 2**. **«ВВЕДЕНИЕ В СТАТИСТИЧЕСКОЕ ОБУЧЕНИЕ»** | |
| Тема 1. Вводные положения | Генеральная совокупность и выборка Типы переменных. Количественные и номинативные переменные. Меры изменчивости. Нормальное распределение. Доверительные интервалы для среднего. Идея статистического вывода, p-уровень значимости. |
| Тема 2. Сравнение средних | Сравнение двух средних, t-критерий. Проверка распределения на нормальность. Однофакторный дисперсионный анализ Множественные сравнения в ANOVAМногофакторный ANOVA |
| Тема 3. Визуальный анализ данных | Квартет Энксомба. Типы визуальных диаграмм и условия их применения. Правила визуализации данных. Типичные ошибки и манипуляции с помощью визуализации. |
| Тема 4. Корреляция и регрессия | Условия применения коэффициента корреляции. Регрессия с одной независимой переменной. Гипотеза о значимости взаимосвязи и коэффициент детерминации. Условия применения линейной регрессии с одним предиктором. Применение регрессионного анализа и интерпретация результатов. Задача предсказания значений зависимой переменной. Регрессионный анализ с несколькими независимыми переменными. Выбор наилучшей модели. |
| **В том числе:** | |
| Практические занятия (семинары, лабораторные занятия) | Планирование статистического эксперимента |
| Анализ мощности выборки |
| Типичные проблемы статистических экспериментов. Поправка Бонферрони |
| Однофакторный дисперсионный анализ |
| Множественные сравнения в ANOVA |
| Многофакторный ANOVA |
| Визуализация данных |
| Регрессия с одной независимой переменной |
| Регрессионный анализ с несколькими независимыми переменными |
| Выбор наилучшей регрессионной модели |
| Регуляризация коэффициентов регрессии и отбор информативных признаков |
| Применяемые образовательные технологии | Электронный курс **https://exam1.urfu.ru/course/view.php?id=844** |
| Рекомендуемые методические материалы, Интернет-ресурсы | Материалы электронного курса https://exam1.urfu.ru/course/view.php?id=844  Электронные информационные ресурсы:   1. Open AI. 2018. How AI Training Scales, [online] Available at: <<https://openai.com/blog/scieiice-of-ai/>>   Учебно-методические материалы:  1. Басараб, М.А. Интеллектуальные технологии на основе искусственных нейронных сетей [Электронный ресурс]: учебное пособие / М.А. Басараб, Н.С. Коннова. - Электрон, дан. - Москва: МГТУ им. Н.Э. Баумана, 2017. - 56 с. - Режим доступа: <https://e.lanbook.com/book/103496>. - Загл. с экрана.  2. Инженерия знаний. Модели и методы [Электронный ресурс] : учебник /Гаврилова Т. А., Кудрявцев Д. В., Муромцев Д. И., 2-е изд., стер. - :Лань, 2018 - 324 с. - Книга из коллекции Лань - Информатика. – Доступ по логину и паролю из сети Интернет .- ISBN 978-5-8114-2128-2 .- <URL:<https://e.lanbook.com/book/107925>>.  3. Кратович, П. В. Нейронные сети и модели ARIMA для прогнозирования котировок [Текст] / П. В. Кратович // Программные продукты и системы. - 2011. - N 1. - С. 95-98: ил.: 3 рис., 3 табл. - ISSN 0236-235Х. - Библиогр.: с. 98 (3 назв.).  4. Малыгин. И. Г. Информационные технологии и искусственный интеллект- основные двигатели четвертой индустриальной революции (Industrie 4.0) = Information Technology and Artificial Intelligence - the Main Engines of the Fourth Industrial Revolution (Industrie 4.0) / И. Г. Малыгин, В. И. Комашинский // Информационные технологии. - 2016. - Т. 22, № 12. - С. 899-904. - (Интеллектуальные системы и технологии). - ISSN 1684-6400. - Библиогр.: с. 904 (13 назв.).  5.Васильев, В. И. Искусственный интеллект в системах управления и обработки информации / В. И. Васильев // Вестник Уфимского государственного авиационного технического университета. - 2000,- №1.- С. 133-140 .- (Научные статьи и доклады. Гуманитарные и социальные науки).- ISSN 1992-6502. - Библиогр.: с. 139-140 (26 назв.). |
| **Модуль 3. «ИНТЕЛЛЕКТУАЛЬНЫЙ АНАЛИЗ ДАННЫХ И МАШИННОЕ ОБУЧЕНИЕ»** | |
| Тема 1. Вводные положения | Понятие интеллектуального анализа данных (DM). Требования, предъявляемые к новым знаниям. Задачи Data Mining. Стадии Data Mining. Обзор типовых задача Data Mining: классификация, кластеризация, поиск ассоциативных правил, регрессионные модели и т. д. Сферы применения технологии интеллектуального анализа данных. Методика обнаружения нового знания в хранилищах и базах данных (KDD). |
| Тема 2. Алгоритмы классификации | Постановка задачи классификации и представление результатов. Классификационные правила: деревья решений; методы построения деревьев решений, ансамбли классификационных алгоритмов, бэггинг и бустинг, случайный лес. логистическая регрессия, наивный байесов классификатор, машина опорных векторов (SVM), метрики качества классификации, проблема переобучения. |
| Тема 3. Методы кластеризации данных и ассоциативные правила | Основные понятия ассоциативных правил, алгоритм Apriori. постановка задачи кластеризации данных, процесс кластеризации данных, представление результатов кластеризации. Агломеративные и дивизимные алгоритмы, алгоритм ближайшего соседа, представление результатов иерархической кластеризации. Итеративные алгоритмы, алгоритм k-means, метрика качества алгоритмов кластеризации, методы определения вероятного числа кластеров, методы визуализации кластерной структуры |
| Тема 4. Методы прогнозирования численных признаков | Регрессионные деревья. Продвинутые регрессионные модели. Прогнозирование временных рядов методами машинного обучения. |
| Тема 5. Факторный анализ и сокращение размерности | Постановка задачи, подготовка к факторному анализу, выделение первичных факторов, метод главных компонент, алгоритм NIPALS, аналитическое вращение факторов, ортогональное и косоугольное вращение, критерии вращения. Методы отбора переменных на основе информационных критериев |
| **В том числе:** | |
| Практические занятия (семинары, лабораторные занятия) | Жизненный цикл аналитики по методологии CRISP-DM |
| Очистка и предобработка данных |
| Классификация с помощью деревьев решений |
| Классификация с помощью метода naïve bayes |
| Кластеризация с помощью алгоритма k-means |
| Прогнозирование временных рядов |
| Факторный анализ |
| Применяемые образовательные технологии | Электронный курс **https://exam1.urfu.ru/course/view.php?id=844** |
| Рекомендуемые методические материалы, Интернет-ресурсы | Материалы электронного курса https://exam1.urfu.ru/course/view.php?id=844  Электронные информационные ресурсы:   1. Open AI. 2018. How AI Training Scales, [online] Available at: <<https://openai.com/blog/scieiice-of-ai/>>   Учебно-методические материалы:  1. Басараб, М.А. Интеллектуальные технологии на основе искусственных нейронных сетей [Электронный ресурс]: учебное пособие / М.А. Басараб, Н.С. Коннова. - Электрон, дан. - Москва: МГТУ им. Н.Э. Баумана, 2017. - 56 с. - Режим доступа: <https://e.lanbook.com/book/103496>. - Загл. с экрана.  2. Инженерия знаний. Модели и методы [Электронный ресурс] : учебник /Гаврилова Т. А., Кудрявцев Д. В., Муромцев Д. И., 2-е изд., стер. - :Лань, 2018 - 324 с. - Книга из коллекции Лань - Информатика. – Доступ по логину и паролю из сети Интернет .- ISBN 978-5-8114-2128-2 .- <URL:<https://e.lanbook.com/book/107925>>.  3. Кратович, П. В. Нейронные сети и модели ARIMA для прогнозирования котировок [Текст] / П. В. Кратович // Программные продукты и системы. - 2011. - N 1. - С. 95-98: ил.: 3 рис., 3 табл. - ISSN 0236-235Х. - Библиогр.: с. 98 (3 назв.).  4. Малыгин. И. Г. Информационные технологии и искусственный интеллект- основные двигатели четвертой индустриальной революции (Industrie 4.0) = Information Technology and Artificial Intelligence - the Main Engines of the Fourth Industrial Revolution (Industrie 4.0) / И. Г. Малыгин, В. И. Комашинский // Информационные технологии. - 2016. - Т. 22, № 12. - С. 899-904. - (Интеллектуальные системы и технологии). - ISSN 1684-6400. - Библиогр.: с. 904 (13 назв.).  5.Васильев, В. И. Искусственный интеллект в системах управления и обработки информации / В. И. Васильев // Вестник Уфимского государственного авиационного технического университета. - 2000,- №1.- С. 133-140 .- (Научные статьи и доклады. Гуманитарные и социальные науки).- ISSN 1992-6502. - Библиогр.: с. 139-140 (26 назв.). |
| **Модуль 4. «ВОПРОСЫ ИСКУСТВЕННОГО ИНТЕЛЛЕКТА »** | |
| Тема 1. Введение в искусственный интеллект | Философские основы искусственного интеллекта. Автономный интеллект. Интеллектуальные агенты. История искусственного интеллекта. Философские проблемы искусственного интеллекта. Технологическая сингулярность. Василиск Роко. |
| Тема 2. Обработка естественного языка | Формальный и статистический подходы к обработке естественного языка. Статистические методы: языковая модель, скрытая марковская модель, алгоритм Витерби, основы лингвистики (токенизация, морфология, выделение частей речи). Парсинг (анализ интересов и зависимостей). |
| Тема 3. Представление знаний | Семантические сети, элементы глубокого обучения для систем основанных на знаниях, онтологии и графы знаний. |
| **В том числе:** | |
| Практические занятия (семинары, лабораторные занятия) | Обработка естественного языка: модель мешка слов для классификации |
| Обработка естественного языка: тематическое моделирование |
| Обработка естественного языка: анализ эмоциональной окраски |
| Обработка естественного языка: скрытые марковские модели и поля |
| Построение графа знаний |
| Применяемые образовательные технологии | Электронный курс **https://exam1.urfu.ru/course/view.php?id=844** |
| Рекомендуемые методические материалы, Интернет-ресурсы | Материалы электронного курса https://exam1.urfu.ru/course/view.php?id=844  Электронные информационные ресурсы:   1. Open AI. 2018. How AI Training Scales, [online] Available at: <<https://openai.com/blog/scieiice-of-ai/>>   Учебно-методические материалы:  1. Басараб, М.А. Интеллектуальные технологии на основе искусственных нейронных сетей [Электронный ресурс]: учебное пособие / М.А. Басараб, Н.С. Коннова. - Электрон, дан. - Москва: МГТУ им. Н.Э. Баумана, 2017. - 56 с. - Режим доступа: <https://e.lanbook.com/book/103496>. - Загл. с экрана.  2. Инженерия знаний. Модели и методы [Электронный ресурс] : учебник /Гаврилова Т. А., Кудрявцев Д. В., Муромцев Д. И., 2-е изд., стер. - :Лань, 2018 - 324 с. - Книга из коллекции Лань - Информатика. – Доступ по логину и паролю из сети Интернет .- ISBN 978-5-8114-2128-2 .- <URL:<https://e.lanbook.com/book/107925>>.  3. Кратович, П. В. Нейронные сети и модели ARIMA для прогнозирования котировок [Текст] / П. В. Кратович // Программные продукты и системы. - 2011. - N 1. - С. 95-98: ил.: 3 рис., 3 табл. - ISSN 0236-235Х. - Библиогр.: с. 98 (3 назв.).  4. Малыгин. И. Г. Информационные технологии и искусственный интеллект- основные двигатели четвертой индустриальной революции (Industrie 4.0) = Information Technology and Artificial Intelligence - the Main Engines of the Fourth Industrial Revolution (Industrie 4.0) / И. Г. Малыгин, В. И. Комашинский // Информационные технологии. - 2016. - Т. 22, № 12. - С. 899-904. - (Интеллектуальные системы и технологии). - ISSN 1684-6400. - Библиогр.: с. 904 (13 назв.).  5.Васильев, В. И. Искусственный интеллект в системах управления и обработки информации / В. И. Васильев // Вестник Уфимского государственного авиационного технического университета. - 2000,- №1.- С. 133-140 .- (Научные статьи и доклады. Гуманитарные и социальные науки).- ISSN 1992-6502. - Библиогр.: с. 139-140 (26 назв.). |

7. КОНТРОЛЬ И ОЦЕНКА РЕЗУЛЬТАТОВ ОСВОЕНИЯ ДОПОЛНИТЕЛЬНОЙ ПРОФЕССИОНАЛЬНОЙ ПРОГРАММЫ

ПОВЫШЕНИЯ КВАЛИФИКАЦИИ

7.1. Система контроля и оценивания

Таблица 3

|  |  |  |
| --- | --- | --- |
| **Результаты**  **(освоенные компетенции)** | **Основные показатели оценки результата** | **Формы и методы контроля** |
| ПК 1) Организовывать работу в соответствии с концепцией жизненного цикла аналитических проектов; | **Знания**  Обучающийся демонстрирует знание-знакомство, знание-копию: узнает объекты, явления и понятия, находит в них различия, проявляет знание источников получения информации, может осуществлять самостоятельно репродуктивные действия над знаниями путем самостоятельного воспроизведения и применения информации.  **Умения**  Обучающийся умеет корректно выполнять предписанные действия по инструкции, алгоритму в известной ситуации, самостоятельно выполняет действия по решению типовых задач, требующих выбора из числа известных методов, в предсказуемо изменяющейся ситуации  **Навыки**  Умеет применять навык разработки и осуществления проектов в соответствии с концепцией жизненного цикла аналитически проектов | Тест, самостоятельное задание |
| ПК 2) Осуществлять процесс ETL (Extraction, Transformation, Loading) из различных видов источников данных | **Знания**  Демонстрирует знание-знакомство, знание-копию: узнает объекты, явления и понятия, находит в них различия, проявляет знание источников получения информации, может осуществлять самостоятельно репродуктивные действия над знаниями путем самостоятельного воспроизведения и применения информации.  **Умения**  Умеет корректно выполнять предписанные действия по инструкции, алгоритму в известной ситуации, самостоятельно выполняет действия по решению типовых задач, требующих выбора из числа известных методов, в предсказуемо изменяющейся ситуации  **Навыки**  Умеет применять навык извлечения и агрегации данных средствами Python | Тест, самостоятельное задание |
| ПК 3) Преобразовывать бизнес-задачи в аналитические проекты | **Знания**  Демонстрирует знание-знакомство, знание-копию: узнает объекты, явления и понятия, находит в них различия, проявляет знание источников получения информации, может осуществлять самостоятельно репродуктивные действия над знаниями путем самостоятельного воспроизведения и применения информации.  **Умения**  Умеет корректно выполнять предписанные действия по инструкции, алгоритму в известной ситуации, самостоятельно выполняет действия по решению типовых задач, требующих выбора из числа известных методов, в предсказуемо изменяющейся ситуации  **Навыки**  Умеет применять навык разработки работы с требованиями в аналитическом проекте | Тест, самостоятельное задание |
| ПК 4) Строить статистические модели и определять их качество. | **Знания**  Демонстрирует знание-знакомство, знание-копию: узнает объекты, явления и понятия, находит в них различия, проявляет знание источников получения информации, может осуществлять самостоятельно репродуктивные действия над знаниями путем самостоятельного воспроизведения и применения информации.  **Умения**  Умеет корректно выполнять предписанные действия по инструкции, алгоритму в известной ситуации, самостоятельно выполняет действия по решению типовых задач, требующих выбора из числа известных методов, в предсказуемо изменяющейся ситуации  **Навыки**  Умеет применять параметрические и непараметрические тестов в статистике, методы корреляции и регрессии | Тест, самостоятельное задание |
| ПК 5) Использовать методы интеллектуального анализа данных и машинного обучения | **Знания**  Демонстрирует знание-знакомство, знание-копию: узнает объекты, явления и понятия, находит в них различия, проявляет знание источников получения информации, может осуществлять самостоятельно репродуктивные действия над знаниями путем самостоятельного воспроизведения и применения информации.  **Умения**  Умеет корректно выполнять предписанные действия по инструкции, алгоритму в известной ситуации, самостоятельно выполняет действия по решению типовых задач, требующих выбора из числа известных методов, в предсказуемо изменяющейся ситуации  **Навыки**  Умеет применять навыки использования методов машинного обучения и интеллектуального анализа данных для решения аналитических задач (кластеризация, классификация и т.п.) | Тест, самостоятельное задание |

7.2. Форма итоговой аттестации – зачет в тестовой форме.

7.3. Примерные задания для итоговой аттестации (темы итоговых работ)

**Примеры тестовых заданий**

1. Какая команда позволяет изменить тип данных в колонке Pandas датафрейма

1. astype() \*\*
2. set()
3. dtype()
4. shape()

2. Какая команда Pandas позволяет узнать структуру DataFrame или Series?

1. astype()
2. set()
3. dtype()
4. shape()\*\*

3. Какая команда Pandas удаляет строки или столбцы содержащие пропуски данных?

1. dropna()\*\*
2. fillna()
3. replace()
4. split()

|  |
| --- |
| 4. Сопоставьте тезисы:   1. Нулевая гипотеза (H0) верна, и вы не смогли принять нулевую гипотезу 2. Нулевая гипотеза (H0) неверна, и вы не смогли отклонить нулевую гипотезу 3. – Ошибка первого типа 4. – Ошибка второго типа |
| 5. Рассмотрим ситуацию:  Предположим, каждый, кто посещает наш сайт, либо получает одно из двух рекламных предложений, либо вообще не получает никакой рекламы.  Мы хотим увидеть, имеет ли значение рекламное предложение.  Какой статистический метод вы бы порекомендовали для этого анализа?   1. P-test 2. T- test 3. Z-test 4. ANOVA \*\* |
| 6. Каково предположение в t-тесте Стьюдента?   1. Выборки нормально распределены\*\* 2. Выборки не нормально распределены 3. Размер выборок мал 4. Не делает никаких предположений относительно выборок |
| 7. Вы анализируете две нормально распределенные выборки, и ваша нулевая гипотеза состоит в том, что среднее m1 первой выборки равно среднему m2 второй. Вы наблюдали значение «p» равное 4,33e-05. Каким будет ваше решение относительно нулевой гипотезой? |

1. p-значение мало, отклоните нулевую гипотезу и примите альтернативную гипотезу \*\*
2. p-значение мало, принимаем нулевую гипотезу и отклоняем альтернативную гипотезу
3. p-Value может не предоставлять достаточной информации и учитывать другую статистику теста
4. Вы должны преобразовать данные с помощью log () и повторить анализ.

8. СОСТАВ ПРЕПОДАВАТЕЛЕЙ

участвующих в реализации программы «Имитационное моделирование организации производственных процессов промышленных предприятий»

Таблица 4

|  |  |  |  |
| --- | --- | --- | --- |
| **№** | **ФИО** | **Место работы, должность, ученое звание** | **Модули/темы** |
|  | Комоцкий Евгений Игоревич | Институт радиоэлектроники и информационных технологий - РТФ УрФУ, ассистент кафедры  "Аналитика больших данных и методы видеоанализа " | Модуль 1-4 |