1. **Паспорт Образовательной программы**

**«** Методы анализа данных и машинного обучения **»**

|  |  |
| --- | --- |
| **Версия программы** | **1** |
| **Дата Версии** | **09.10.2020** |

1. **Сведения о Провайдере**

|  |  |  |
| --- | --- | --- |
| 1.1 | Провайдер | ФГБОУ ВО "Липецкий государственный технический университет" |
| 1.2 | Логотип образовательной организации |  |
| 1.3 | Провайдер ИНН | 4826012416 |
| 1.4 | Ответственный за программу ФИО | Корнеева Анастасия Олеговна |
| 1.5 | Ответственный должность | Специалист по учебно-методической работе деканата факультета дополнительного образования |
| 1.6 | Ответственный Телефон | +7(904)285-23-83 |
| 1.7 | Ответственный Е-mail | korneeva\_ao@stu.lipetsk.ru |

1. **Основные Данные**

|  |  |  |
| --- | --- | --- |
| **№** | **Название** | **Описание** |
| 2.1 | Название программы | Методы анализа данных и машинного обучения |
| 2.2 | Ссылка на страницу программы | <http://eserv.stu.lipetsk.ru/course/index.php?categoryid=37>  <https://cloud.stu.lipetsk.ru/index.php/s/qXByQiKztaY4dBJ#pdfviewer> |
| 2.3 | Формат обучения | Онлайн |
|  | Подтверждение от ОО наличия возможности реализации образовательной программы с применением электронного обучения и (или) дистанционных образовательных технологий с возможностью передачи данных в форме элементов цифрового следа | У каждого обучающегося будет заведен личный кабинет на сайте ЛГТУ.  <http://eserv.stu.lipetsk.ru/course/view.php?id=663> |
| 2.4 | Уровень сложности | Продвинутый |
| 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) | 64 |
| 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 | 30 000 руб.  МФТИ (Машинное обучение и анализ данных) <https://mipt.ru/cdpo/programs/software/supervised_learning.php?ELEMENT_ID=1855780> 150 ч. 69600 руб. ВШЭ (Основы машинного обучения и продвинутые методы машинного обучения) <https://cs.hse.ru/dpo/advancedml> 114 ч. 88000 руб. Финансовый университет (Искусственный интеллект, большие данные и машинное обучение) <http://xn--80aaoligqe4brp0a.xn--p1ai/programms/iibd/> (краткосрочная программа) 19800 руб. |
| 2.7 | Минимальное количество человек на курсе | 10 |
| 2.8 | Максимальное количество человек на курсе | 45 |
| 2.9 | Данные о количестве слушателей, ранее успешно прошедших обучение по образовательной программе | При наличии |
| 2.10 | Формы аттестации | Итоговое задание |
|  | 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**

Наиболее полное и содержательное описание программы, которое включает:

1) общую характеристику компетенций, качественное изменение которых осуществляется в результате обучения или которые формируются в результате освоения образовательной программы;

2) описание требований и рекомендаций для обучения по образовательной программе;

3) краткое описание результатов обучения в свободной форме, а также описание востребованности результатов обучения в профессиональной деятельности.

Ограничение по размеру: не менее 1000 символов

1. Цель реализации программы

Целью реализации программы является формирование представления о современных методах анализа данных различной природы, а также об использовании моделей машинного обучения для решения задач прикладного характера.

2. Требования к результатам обучения

В результате освоения программы слушатель должен приобрести следующие знания:

основные компьютерные технологии проведения прикладных математических ис-следований;

специализированные пакеты программ, предназначенные для решения прикладных математических задач;

В результате освоения программы слушатель должен приобрести следующие умения:

проводить экспериментальные научные исследования различных явлений;

выбирать нужные математические модели и методы исследования, а также соответствующие им программные средства и профессионально использовать компьютерные технологии для решения прикладных задач;

составлять алгоритмы с учётом специфики машинных вычислений и программировать на языке инженерных и научных расчетов R.

В результате освоения программы слушатель должен приобрести следующие навыки:

использование средств обработки информации;

работа с современным программным обеспечением для математических исследований.

ФГБОУ ВО «Липецкий государственный технический университет»

УТВЕРЖДАЮ

Ректор ЛГТУ

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ Сараев П.В.

« \_\_\_ » \_\_\_\_\_\_\_\_\_\_\_\_\_ 2020 г.

Дополнительная профессиональная программа повышения квалификации

« Методы анализа данных и машинного обучения »

72 часа.

**ОБЩАЯ ХАРАКТЕРИСТИКА ПРОГРАММЫ**

**1.Цель программы**

Целью реализации программы является формирование представления о современных методах анализа данных различной природы, а также об использовании моделей машинного обучения для решения задач прикладного характера.

**2.Планируемые результаты обучения:**

2.1.Знание (осведомленность в областях)

2.1.1. Основные компьютерные технологии проведения прикладных математических исследований;

2.1.2. Специализированные пакеты программ, предназначенные для решения прикладных математических задач.

2.2. Умение (способность к деятельности)

2.2.1. Проводить экспериментальные научные исследования различных явлений;

2.2.2. Выбирать нужные математические модели и методы исследования, а также соответствующие им программные средства и профессионально использовать компьютерные технологии для решения прикладных задач;

2.2.3. Составлять алгоритмы с учётом специфики машинных вычислений и программировать на языке инженерных и научных расчетов R.

2.3.Навыки (использование конкретных инструментов)

2.3.1 Использование средств обработки информации;

2.3.2. Работа с современным программным обеспечением для математических исследований.

**3.Категория слушателей** (возможно заполнение не всех полей)

* 1. Среднее специальное / высшее техническое / высшее экономическое
  2. Не имеет значения
  3. Не имеет значения
  4. Не имеет значения

**4.Учебный план программы «…..наименование программы….»**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **№ п/п** | **Модуль** | **Всего, час** | **Виды учебных занятий** | | |
| **лекции** | **практические занятия** | **самостоятельная работа** |
| 1 | Разведочный анализ и визуализация данных | 10 | 2 | 2 | 6 |
| 2 | Методы анализа структуры данных | 10 | 2 | 2 | 6 |
| 3 | Анализ взаимосвязей – корреляционный анализ. Статистические гипотезы. Проверка гипотезы о значимости коэффициента корреляции | 6 | 2 | 2 | 2 |
| 4 | Установление структуры связи – линейная регрессионная модель | 12 | 4 | 4 | 4 |
| 5 | Задача разделения наблюдений на группы без обучения – кластеризация | 12 | 4 | 4 | 4 |
| 6 | Искусственные нейронные сети | 14 | 4 | 4 | 6 |
| **Итоговая аттестация** | |  | **Указывается вид (экзамен, зачёт, реферат и т.д.)** | | |
| Итоговый контроль | | 8 | Выполнение индивидуального задания | | |

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

(дата начала обучения – дата завершения обучения) в текущем календарном году, указания на периодичность набора групп (не менее 1 группы в месяц)

|  |  |  |  |
| --- | --- | --- | --- |
| **№ п/п** | **Наименование учебных модулей** | **Трудоёмкость (час)** | **Сроки обучения** |
| **1** | Разведочный анализ и визуализация данных | 10 | 18.11.2020 |
| **2** | Методы анализа структуры данных | 10 | 19.11.2020 |
| 3 | Анализ взаимосвязей – корреляционный анализ. Статистические гипотезы. Проверка гипотезы о значимости коэффициента корреляции | 6 | 20.11.2020 |
| 4 | Установление структуры связи – линейная регрессионная модель | 12 | 23.11.2020 |
| 5 | Задача разделения наблюдений на группы без обучения – кластеризация | 12 | 24.11.2020 |
| 6 | Искусственные нейронные сети | 14 | 25.11.2020 |
| **Всего:** | | 72 | 18-25.11.2020 г. |

**6.Учебно-тематический план программы «** Методы анализа данных и машинного обучения **»**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **№ п/п** | **Модуль / Тема** | **Всего, час** | **Виды учебных занятий** | | | **Формы контроля** |
| **лекции** | **практические занятия** | **самостоятельная работа** |
| 1 | Разведочный анализ и визуализация данных | 10 | 2 | 2 | 6 | Проверка заданий, выполненных в рамках самостоятельной работы |
| 1.1 | Источники данных. Генеральная совокупность и выборка | 1 | 1 | 0 | 0 | - |
| 1.2 | Типы данных в языке R. Векторы и матрицы. Факторы. Списки  и таблицы | 2 | 1 | 0 | 1 | Проверка заданий, выполненных в рамках самостоятельной работы |
| 1.3 | Библиотеки R для импортирования данных. Импорт данных из текстовых файлов. Импорт из таблиц MS Excel | 1,5 | 0 | 0,5 | 1 | Проверка заданий, выполненных в рамках самостоятельной работы |
| 1.4 | Сортировка данных. Разделение наборов данных на составляющие | 0,5 | 0 | 0 | 0,5 | Проверка заданий, выполненных в рамках самостоятельной работы |
| 1.5 | Графические методы первичного анализа данных. Гистограммы. Диаграммы размахов. Круговые и столбиковые диаграммы | 0,5 | 0 | 0 | 0,5 | Проверка заданий, выполненных в рамках самостоятельной работы |
| 1.6 | Базовые графические возможности R. Представление двумерных данных. Функция plot() и ее аргументы | 1 | 0 | 0,5 | 0,5 | Проверка заданий, выполненных в рамках самостоятельной работы |
| 1.7 | Представление многомерных данных. Диаграммы рассеяния. Пиктограммы. Графики параллельных координат | 1 | 0 | 0,5 | 0,5 | Проверка заданий, выполненных в рамках самостоятельной работы |
| 1.8 | Представление многомерных данных в R. Пакеты rgl, lattice | 1 | 0 | 0,5 | 0,5 | Проверка заданий, выполненных в рамках самостоятельной работы |
| 1.9 | Примеры решения практических задач графического анализа данных | 1,5 | 0 | 0 | 1,5 | Проверка заданий, выполненных в рамках самостоятельной работы |
| 2 | Методы анализа структуры данных | 10 | 2 | 2 | 6 | Проверка заданий, выполненных в рамках самостоятельной работы |
| 2.1 | Описательная статистика. Применение функции summary | 0,5 | 0,5 | 0 | 0 | - |
| 2.2 | Устранение пропущенных данных. Нахождение ошибочных наблюдений. | 1,5 | 0,5 | 0 | 1 | Проверка заданий, выполненных в рамках самостоятельной работы |
| 2.3 | Обнаружение пропущенных значений. Исследование структуры пропущенных данных | 1,5 | 0 | 0,5 | 1 | Проверка заданий, выполненных в рамках самостоятельной работы |
| 2.4 | Пакеты mice и VIM для исследования пропусков. | 1,5 | 0 | 0,5 | 1 | Проверка заданий, выполненных в рамках самостоятельной работы |
| 2.5 | Восстановления пропущенных данных. Пакет Amelia | 1,5 | 0 | 0,5 | 1 | Проверка заданий, выполненных в рамках самостоятельной работы |
| 2.6 | Необычные наблюдения. Выбросы. Точки высокой напряженности. Влиятельные наблюдения. | 1,5 | 0,5 | 0 | 1 | Проверка заданий, выполненных в рамках самостоятельной работы |
| 2.7 | Методы R для обнаружения необычных наблюдений. | 0,5 | 0,5 | 0 | 0 | - |
| 2.8 | Примеры решения задач исследования структуры данных, выявления необычных наблюдений, восстановления пропусков. | 1,5 | 0 | 0,5 | 1 | Проверка заданий, выполненных в рамках самостоятельной работы |
| 3 | Анализ взаимосвязей – корреляционный анализ. Статистические гипотезы. Проверка гипотезы о значимости коэффициента корреляции. | 6 | 2 | 2 | 2 | Проверка заданий, выполненных в рамках самостоятельной работы |
| 4 | Установление структуры связи – линейная регрессионная модель | 12 | 4 | 4 | 4 | Проверка заданий, выполненных в рамках самостоятельной работы |
| 4.1 | Построение регрессионной модели в R. Функция lm и ее аргументы. Проверка гипотез о значимости регрессионных моделей. | 2 | 1 | 1 | 0 | - |
| 4.2 | Диагностика регрессионных моделей. Способы корректировки регрессионной модели | 3 | 1 | 1 | 1 | Проверка заданий, выполненных в рамках самостоятельной работы |
| 4.3 | Множественная линейная регрессия со взаимодействиями | 3 | 1 | 1 | 1 | Проверка заданий, выполненных в рамках самостоятельной работы |
| 4.4 | Сравнение моделей и выбор лучшей | 2 | 1 | 0 | 1 | Проверка заданий, выполненных в рамках самостоятельной работы |
| 4.5 | Практические задачи моделирования состояния процессов и систем с использованием методов регрессионного анализа | 2 | 0 | 1 | 1 | Проверка заданий, выполненных в рамках самостоятельной работы |
| 5 | Задача разделения наблюдений на группы без обучения – кластеризация | 12 | 4 | 4 | 4 | Проверка заданий, выполненных в рамках самостоятельной работы |
| 5.1 | Задача кластеризации. Меры расстояния. Пакет cluster | 3 | 1 | 1 | 1 | Проверка заданий, выполненных в рамках самостоятельной работы |
| 5.2 | Кластеризация методом k-средних. Функция kmeans и ее аргументы. | 3 | 1 | 1 | 1 | Проверка заданий, выполненных в рамках самостоятельной работы |
| 5.3 | Интерпретация результатов кластеризации. Определение рационального числа кластеров. Иерархическая кластеризация. | 3 | 1 | 1 | 1 | Проверка заданий, выполненных в рамках самостоятельной работы |
| 5.4 | Примеры решения задач кластеризации данных различной структуры. | 3 | 1 | 1 | 1 | Проверка заданий, выполненных в рамках самостоятельной работы |
| 6 | Искусственные нейронные сети | 14 | 4 | 4 | 6 | Проверка заданий, выполненных в рамках самостоятельной работы |
| 6.1 | Структура нейросетевой модели. Принцип работы нейронной сети. Функция активации. Параметры и гиперпараметры нейронной сети. | 1 | 1 | 0 | 0 | - |
| 6.2 | Библиотека neuralnet и ее функции. | 0,5 | 0 | 0,5 | 0 | - |
| 6.3 | Обработка данных для обучения модели и для дальнейшего прогноза. Функция scale. | 1,5 | 0 | 0,5 | 1 | Проверка заданий, выполненных в рамках самостоятельной работы |
| 6.4 | Обучение нейросетевой модели. Функция neuralnet и ее параметры. | 2,5 | 1 | 0,5 | 1 | Проверка заданий, выполненных в рамках самостоятельной работы |
| 6.5 | Оценка качества нейросетевой модели. Меры ошибок. | 2 | 1 | 0 | 1 | Проверка заданий, выполненных в рамках самостоятельной работы |
| 6.6 | Прогнозирование с использованием полученной модели. Функция compute и ее аргументы | 2,5 | 1 | 0,5 | 1 | Проверка заданий, выполненных в рамках самостоятельной работы |
| 6.7 | Решение практических задач прогнозирования с использованием нейронных сетей. | 2 | 0 | 1 | 1 | Проверка заданий, выполненных в рамках самостоятельной работы |
| 6.8 | Решение практических задач нейросетевой классификации. | 2 | 0 | 1 | 1 | Проверка заданий, выполненных в рамках самостоятельной работы |

**7. Учебная (рабочая) программа повышения квалификации «**  **название**   **»**

Модуль 1. Разведочный анализ и визуализация данных (10 ч.)

Тема 1.1. Источники данных. Генеральная совокупность и выборка (1 ч.)

Тема  1.2. Типы данных в языке R. Векторы и матрицы. Факторы. Списки и таблицы (2 ч.)

Тема  1.3. Библиотеки R для импортирования данных. Импорт данных из текстовых файлов. Импорт из таблиц MS Excel (1,5 ч.)

Тема  1.4. Сортировка данных. Разделение наборов данных на составляющие (0,5 ч.)

Тема  1.5. Графические методы первичного анализа данных. Гистограммы. Диаграммы размахов. Круговые и столбиковые диаграммы (1 ч.)

Тема  1.6. Базовые графические возможности R. Представление двумерных данных. Функция plot() и ее аргументы (0,5 ч.)

Тема  1.7. Представление многомерных данных. Диаграммы рассеяния. Пиктограммы. Графики параллельных координат (1 ч.)

Тема  1.8. Представление многомерных данных в R. Пакеты rgl, lattice (1 ч.)

Тема  1.9. Примеры решения практических задач графического анализа данных (1,5 ч.)

Модуль 2. Методы анализа структуры данных (10 ч.)

Тема  2.1. Описательная статистика. Применение функции summary (0,5 ч.)

Тема  2.2. Устранение пропущенных данных. Нахождение ошибочных наблюдений (1,5 ч.)

Тема  2.3. Обнаружение пропущенных значений. Исследование структуры пропущенных данных (1,5 ч.)

Тема  2.4. Пакеты mice и VIM для исследования пропусков (1,5 ч.)

Тема  2.5. Восстановления пропущенных данных. Пакет Amelia (1,5 ч.)

Тема  2.6. Необычные наблюдения. Выбросы. Точки высокой напряженности. Влиятельные наблюдения (1,5 ч.)

Тема  2.7. Методы R для обнаружения необычных наблюдений (0,5 ч.)

Тема  2.8. Примеры решения задач исследования структуры данных, выявления необычных наблюдений, восстановления пропусков (1,5 ч.)

Модуль 3. Анализ взаимосвязей – корреляционный анализ. Статистические гипотезы. Проверка гипотезы о значимости коэффициента корреляции (6 ч.)

Модуль 4. Установление структуры связи – линейная регрессионная модель (12 ч.)

Тема  4.1. Построение регрессионной модели в R. Функция lm и ее аргументы. Проверка гипотез о значимости регрессионных моделей (2 ч.)

Тема  4.2. Диагностика регрессионных моделей. Способы корректировки регрессионной модели (3 ч.)

Тема  4.3. Множественная линейная регрессия со взаимодействиями (3 ч.)

Тема  4.4. Сравнение моделей и выбор лучшей (2 ч.)

Тема  4.5. Практические задачи моделирования состояния процессов и систем с использованием методов регрессионного анализа (2 ч.)

Модуль 5. Задача разделения наблюдений на группы без обучения – кластеризация (12 ч.)

Тема  5.1. Задача кластеризации. Меры расстояния. Пакет cluster (3 ч.)

Тема  5.2. Кластеризация методом k-средних. Функция kmeans и ее аргументы (3 ч.)

Тема  5.3. Интерпретация результатов кластеризации. Определение рационального числа кластеров. Иерархическая кластеризация (3 ч.)

Тема  5.4. Примеры решения задач кластеризации данных различной структуры (3 ч.)

Модуль 6. Искусственные нейронные сети (14 ч.)

Тема  6.1. Структура нейросетевой модели. Принцип работы нейронной сети. Функция активации. Параметры и гиперпараметры нейронной сети (1 ч.)

Тема  6.2. Библиотека neuralnet и ее функции (0,5 ч.)

Тема  6.3. Обработка данных для обучения модели и для дальнейшего прогноза. Функция scale (1,5 ч.)

Тема  6.4. Обучение нейросетевой модели. Функция neuralnet и ее параметры (2,5 ч.)

Тема  6.5. Оценка качества нейросетевой модели. Меры ошибок (2 ч.)

Тема  6.6. Прогнозирование с использованием полученной модели. Функция compute и ее аргументы (2,5 ч.)

Тема  6.7. Решение практических задач прогнозирования с использованием нейронных сетей (2 ч.)

Тема  6.8. Решение практических задач нейросетевой классификации (2 ч.)

**Описание практико-ориентированных заданий и кейсов**

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Номер темы/модуля** | **Наименование практического занятия** | **Описание** |
| 1 | 1.7 | Представление многомерных данных. Диаграммы рассеяния. Пиктограммы. Графики параллельных координат | Кейс-задание на формирование и анализ таблицы данных. Построение графического представления многомерных данных. |
| **2** | 1.9 | Примеры решения практических задач графического анализа данных | Кейс-задание на построение графика параллельных координат и решение на его основе первичных задач анализа |
| 3 | 2.8 | Примеры решения задач исследования структуры данных, выявления необычных наблюдений, восстановления пропусков | Кейс-задание на восстановление пропущенных данных в массиве, оценку качества реализованной процедуры |
| 4 | 3 | Анализ взаимосвязей – корреляционный анализ | Кейс-задание на определение степени линейной связи меду переменными в наборе данных |
| 5 | 4.1 | Построение регрессионной модели в R. Функция lm и ее аргументы. Проверка гипотез о значимости регрессионных моделей | Кейс-задание на построение линейной МНК-модели, прогнозирование с ее помощью |
| 6 | 4.5 | Практические задачи моделирования состояния процессов и систем с использованием методов регрессионного анализа | Кейс-задание на построение линейной МНК-модели со взаимодействиями, ее корректировку, прогнозирование с ее помощью |
| 7 | 5.4 | Примеры решения задач кластеризации данных различной структуры | Кейс-задание на методы иерархической кластеризации и кластеризации методом k-средних |
| 8 | 6.7 | Решение практических задач прогнозирования с использованием нейронных сетей | Кейс-задание на построение состоятельной нейросетевой модели прогнозирования уровня качества образования, полученного студентом |

**8.Оценочные материалы по образовательной программе**

**8.1. Вопросы тестирования по модулям**

|  |  |  |  |
| --- | --- | --- | --- |
| **№ модуля** | **Вопросы входного тестирования** | **Вопросы промежуточного тестирования** | **Вопросы итогового тестирования** |
| 1 | - |  | Откуда берутся данные?  Генеральная совокупность и выборка. Как получить данные? Что ищут в дан-ных? Как обрабатывать данные?  Анализ данных. Что это и как его выполнять? Рабочее пространство R. Пакеты. |
| 2 | Типы данных в языке R. Представление даты и времени; временные ряды. Организация вычислений: функции, ветвления, циклы. Базовые графические возможности. | Основные статистические функции.  Стандартизация данных.  Проверка распределения на нормальность.  Необычные наблюдения. | Статистические распределения. Подгонка статистического распределения.  Этапы работы с пропущенными данными. Обнаружение пропущенных значений. Исследование структуры пропущенных данных.  Выявление источников пропущенных данных и эффекта от них. Анализ полных строк (построчное удаление).  Метод множественного восстановления пропущенных данных.  Попарное удаление. |
| 3 | Корреляции. Типы корреляций. | Проверка статистической значимости корреляций. | - |
| 4 | - | Диагностика регрессионных моделей.  Способы корректировки регрессионной модели. | Регрессии. МНК-регрессии.  Сравнение моделей и выбор лучшей. |
| 5 | Расстояние между точками на плоскости и в пространстве | Решение задачи дискретной оптимизации | Походы к кластеризации. Определение числа кластеров. Выбор рационального разбиения реализаций на кластеры |
| 6 | Производная непрерывной функции. Создание массивов с данными. | Параметрическая идентификация нейросетевой модели. | Прогнозирование с использованием нейросетевой модели. Решение задачи классификации с применением обученной нейросетевой структуры. |

**8.2.**  Входной контроль охватывает всех обучаемых и проводится в форме собеседования. Целью его является определение уровня знаний обучаемых. Итоговая аттестация проводится в форме выполнения индивидуальных заданий. Итоговая работа рассчитана на 8 ч., выполняется как кейс-задание в рамках самостоятельной работы, позволяет оценить знания, умения и уровень приобретенных компетенций.

**.**

**8.3.**  Блок «Разведочный анализ и визуализация данных»

Создать датафрейм, содержащий 5 переменных и 10 реализаций, и заполнить каждую переменную случайными значениями от – ni до ni, i = 1,…,5.

Блок «Методы анализа структуры данных»

Задан вектор A <- c (33, 21, 12, NA, 7, 8). Приведите код для подсчета среднего значения компонентов вектора без учета пропущенного значения.

Задан набор данных, содержащих пропущенные значения. Замените в каждой переменной пропущенные значения на медианные значения по соответствующей переменной.

Задан одномерный набор данных с зафиксированными значениями показателя некоторого процесса. С помощью стандартных функций определите необычные наблюдения в этом массиве и создайте новый массив, не содержащий этих наблюдений.

Блок «Анализ взаимосвязей – корреляционный анализ. Статистические гипотезы. Проверка гипотезы о значимости коэффициента корреляции»

Используя стандартную функцию cor и заданный массив данных вычислите матрицу корреляционных коэффициентов. Дайте интерпретацию полученным значениям.

**.**

**8.4.**  Блок «Установление структуры связи – линейная регрессионная модель»

Таблица данных содержит информацию о высоте, обхвате ствола 32 яблонь, а также объеме урожая, полученном с этих деревьев.

Построить модель парной линейной регрессии, где в качестве зависимой переменной выбран объем урожая, а в качестве независимой переменной – высота / обхват ствола.

Оценить значимость параметров полученной модели, вычислить для них доверительные интервалы.

Построить таблицу с данными, где в первом столбце содержатся наблюдаемые, а во втором – прогнозные значения. Оценить остатки модели, при необходимости удалить влиятельные наблюдения для улучшение качества модели.

Построить модель зависимости логарифма объясняемой переменной от логарифма объясняющей переменной. При помощи информационного критерия Акаике сравнить построенную модель с первоначальной.

Построить прогноз сбора урожая для дерева с заданными параметрами.

Блок «Задача разделения наблюдений на группы без обучения – кластеризация»

Задана таблица с данными о 41 европейском городе, содержащая в себе информацию о различных показателях (население, половозрастной состав, географические показателя, экономические показатели и т.п.). Необходимо:

Построить дендрограмму для исходных данных, выдвинуть гипотезу о числе кластеров.

Методом k-средних произвести кластеризацию городов. Дать содержательную интерпретацию полученных результатов.

Повторить пункту a-b для стандартизованных исходных данных. Сравнить полученные результаты.

Блок «Искусственные нейронные сети»

Задана таблица с данными о социальных, экономических, географических показателях, объясняющих качество полученного студентом образования. Необходимо:

Произвести нормализацию данных (привести их к одинаковым безразмерным величинам).

Разделить полученный массив данных на обучающую и тестовую выборки.

С помощью функции neuralnet построить нейросетевую модель зависимости с заданными метапараметрами.

Получить прогноз по построенной модели для тестовой выборки и оценить качество полученной модели. **.**

**8.5.**  Итоговое кейс-задание оценивается в:

1-2 балла - «неудовлетворительно», не зачтено.

3 балла - «удовлетворительно», зачтено.

4 балла - «хорошо», зачтено.

5 баллов - «отлично», зачтено. **.**

**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** | Сараев Павел Викторович | ЛГТУ, ректор, д.т.н., доцент | http://www.stu.lipetsk.ru/struct/management/rector/ |  |  |
| **2** | Сысоев Антон Сергеевич | ЛГТУ, доцент, к.т.н., доцент | https://www.researchgate.net/profile/Anton\_Sysoev2 |  |  |

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

|  |  |
| --- | --- |
| **Учебно-методические материалы** | |
| Методы, формы и технологии | Методические разработки,  материалы курса, учебная литература |
|  | Презентации курса. Скрипты. |
|  | Шипунов А. Б., Балдин E. М., Волкова П. А., Коробейников А. И., Назарова С. А., Петров С. В., Суфиянов В. Г. (2012) Наглядная статистика. Используем R! - М.: ДМК Пресс, 298 с.  Мастицкий С. Э., Шитиков В. К. (2014) Статистический анализ и визуализация данных с помощью R. - Электронная книга, 400 с.  Кабаков Р. R в действии. Анализ и визуализация данных на языке R (2013) - М.: ДМК Пресс, 580 с.  <https://www.manning.com/books/r-in-action-second-edition> |

|  |  |
| --- | --- |
| **Информационное сопровождение** | |
| Электронные  образовательные ресурсы | Электронные  информационные ресурсы |
| Сайт r-analytics | <https://r-analytics.blogspot.com/> |
| Сайт r-exercices | <https://www.r-exercises.com/> |
| Онлайн-компилятор R | https://rextester.com/l/r\_online\_compiler |

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

|  |  |
| --- | --- |
| Вид занятий | Наименование оборудования,  программного обеспечения |
| Практические | ЭВМ с доступом к сети Internet или установленной средой RStudio |
| Самостоятельная работа | ЭВМ с доступом к сети Internet или установленной средой RStudio |

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

Описание перечня профессиональных компетенций, качественное изменение которых осуществляется в результате обучения.

Планируемые результаты обучения должны быть определены в виде знаний, умений, навыков и (или) опыта деятельности, характеризующих формирование/развитие компетенции(-й) в области цифровой экономики и представлены в виде Паспорта компетенций в машиночитаемом текстовом формате. Структура паспорта представлена в приложении.

ПАСПОРТ КОМПЕТЕНЦИИ

дополнительной профессиональной образовательной

программы повышения квалификации

«Методы анализа данных и машинного обучения»

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 1. | Наименование компетенции | | способность к абстрактному мышлению, анализу, синтезу | |
| 2. | Указание типа компетенции | общекультурная/  универсальная | + | |
| общепрофессиональная |  | |
| профессиональная |  | |
| профессионально-специализированная |  | |
| 3. | Определение, содержание и основные сущностные характеристики компетенции | | Умения: проводить экспериментальные научные исследования различных явлений  Навыки: использование средств обработки информации | |
| 4. | Дескриптор знаний, умений и навыков по уровням | | Уровни сформирован ности компетенции  обучающегося | Индикаторы |
| ЗНАТЬ: методы абстрактного мышления при установлении истины, методы научного исследования путём мысленного расчленения объекта (анализ) и путём изучения предмета в его целостности, единстве его частей (синтез) | | Начальный уровень | Общие, но не структурированные знания методов абстрактного мышления, анализа и синтеза при решении исследовательских и практических задач |
| Базовый уровень | Сформированные, но содержащие отдельные пробелы знания основных методов к абстрактного мышления, анализа и синтеза при решении исследовательских и практических задач |
| Профессиональный | Сформированные систематические знания методов абстрактного мышления, анализа и синтеза при решении исследовательских и практических задач |
| УМЕТЬ: с использованием методов абстрактного мышления, анализа и синтеза анализировать альтернативные варианты решения исследовательских задач и оценивать экономическую эффективность реализации этих вариантов | | Начальный уровень | В целом успешно, но не систематически осуществляемые анализ альтернативных вариантов решения исследовательских задач и оценка экономической эффективности реализации этих вариантов |
| Базовый уровень | В целом успешные, но содержащие отдельные пробелы анализ альтернативных вариантов решения исследовательских задач и оценка экономической эффективности реализации этих вариантов |
| Профессиональный | Сформированное умение анализировать альтернативные варианты решения исследовательских задач и оценивать экономическую эффективность реализации этих вариантов |
| ВЛАДЕТЬ: целостной системой навыков использования абстрактного мышления при решении проблем, возникающих при выполнении исследовательских работ, навыками отстаивания своей точки зрения | | Начальный уровень | В целом успешное, но не систематическое применение навыков методологического использования абстрактного мышления при решении проблем, возникающих при выполнении исследовательских работ, самостоятельного мышления, отстаивания своей точки зрения |
| Базовый уровень | В целом успешное, но содержащее отдельные пробелы применение навыков методологического использования абстрактного мышления при решении проблем, возникающих при выполнении исследовательских работ, самостоятельного мышления, отстаивания своей точки зрения |
| Профессиональный | Успешное и систематическое применение навыков методологического использования абстрактного мышления при решении проблем, возникающих при выполнении исследовательских работ, самостоятельного мышления, отстаивания своей точки зрения |
| 5. | Характеристика взаимосвязи данной компетенции с другими компетенциями/ необходимость владения другими компетенциями для формирования данной компетенции | | Успешное формирование компетенции «способность к абстрактному мышлению, анализу, синтезу» служит основой для освоения компетенции «способность самостоятельно приобретать с помощью информационных технологий и использовать в практической деятельности новые знания и умения, в том числе, в новых областях знаний, непосредственно не связанных со сферой деятельности, расширять и углублять свое научное мировоззрение» | |
| 6. | Средства и технологии оценки | | Набор практических заданий для выявления уровня сформированности компетенции | |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 1. | Наименование компетенции | | способность самостоятельно приобретать с помощью информационных технологий и использовать в практической деятельности новые знания и умения, в том числе, в новых областях знаний, непосредственно не связанных со сферой деятельности, расширять и углублять свое научное мировоззрение | |
| 2. | Указание типа компетенции | общекультурная/  универсальная |  | |
| общепрофессиональная | + | |
| профессиональная |  | |
| профессионально-специализированная |  | |
| 3. | Определение, содержание и основные сущностные характеристики компетенции | | Умения: выбирать нужные математические модели и методы исследования, а также соответствующие им программные средства и профессионально использовать компьютерные технологии для решения прикладных задач | |
| 4. | Дескриптор знаний, умений и навыков по уровням | | Уровни сформирован ности компетенции  обучающегося | Индикаторы |
| ЗНАТЬ: знать основные методы и принципы поиска и классификации информации в интернете и электронных библиотеках | | Начальный уровень | Имеет не полные представления об основных методах и принципах поиска и классификации информации в интернете и электронных библиотеках |
| Базовый уровень | Сформированные, но содержащие отдельные пробелы представления об основных методах и принципах поиска и классификации информации в интернете и электронных библиотеках |
| Профессиональный | Отлично знает основные методы и принципы поиска и классификации информации в интернете и электронных библиотеках |
| УМЕТЬ: уметь находить, классифицировать и оценивать найденную информацию, а так же использовать ее для расширения своего научного мировоззрения | | Начальный уровень | Не умеет находить, классифицировать и оценивать найденную информацию, а так же использовать ее для расширения своего научного мировоззрения |
| Базовый уровень | Слабо умеет находить, классифицировать и оценивать найденную информацию, а так же использовать ее для расширения своего научного мировоззрения |
| Профессиональный | Хорошо умеет находить, классифицировать и оценивать найденную информацию, а так же использовать ее для расширения своего научного мировоззрения |
| ВЛАДЕТЬ: навыками самообразования, в том числе - использования интернета и нейросетей в поиске и классификации найденной информации; владеть навыками применения найденной информации для расширения и углубления своего научного мировоззрения | | Начальный уровень | Частично владеет навыками самообразования, в том числе - использования интернета и нейросетей в поиске и классификации найденной информации; владеть навыками применения найденной информации для расширения и углубления своего научного мировоззрения |
| Базовый уровень | Достаточно хорошо владеет навыками самообразования, в том числе - использования интернета и нейросетей в поиске и классификации найденной информации; владеть навыками применения найденной информации для расширения и углубления своего научного мировоззрения |
| Профессиональный | Свободно владеет навыками самообразования, в том числе - использования интернета и нейросетей в поиске и классификации найденной информации; владеть навыками применения найденной информации для расширения и углубления своего научного мировоззрения |
| 5. | Характеристика взаимосвязи данной компетенции с другими компетенциями/ необходимость владения другими компетенциями для формирования данной компетенции | | Успешное формирование компетенции «способность самостоятельно приобретать с помощью информационных технологий и использовать в практической деятельности новые знания и умения, в том числе, в новых областях знаний, непосредственно не связанных со сферой деятельности, расширять и углублять свое научное мировоззрение» базируется на освоении компетенции «способность к абстрактному мышлению, анализу, синтезу» | |
| 6. | Средства и технологии оценки | | Набор практических заданий для выявления уровня сформированности компетенции | |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 1. | Наименование компетенции | | способность анализировать и применять углубленные знания в области прикладной математики и информатики | |
| 2. | Указание типа компетенции | общекультурная/  универсальная |  | |
| общепрофессиональная |  | |
| профессиональная | + | |
| профессионально-специализированная |  | |
| 3. | Определение, содержание и основные сущностные характеристики компетенции | | Знания: специализированные пакеты программ, предназначенные для решения прикладных математических задач  Умения: составлять алгоритмы с учётом специфики машинных вычислений и программировать на языке инженерных и научных расчетов R  Навыки: работа с современным программным обеспечением для математических исследований | |
| 4. | Дескриптор знаний, умений и навыков по уровням | | Уровни сформирован ности компетенции  обучающегося | Индикаторы |
| ЗНАТЬ:  основные принципы и методы разработки алгоритмов | | Начальный уровень | Имеет не полные представления об основных принципах и методах разработки алгоритмов |
| Базовый уровень | Сформированные, но содержащие отдельные пробелы в представлениях об основных принципах и методах разработки алгоритмов |
| Профессиональный | Отлично знает основные принципы и методы разработки алгоритмов |
| УМЕТЬ:  ставить новые задачи в области прикладной математики и информатики и находить пути их решения, формулировать задачи, а так же разрабатывать алгоритмы для написания компьютерных программ | | Начальный уровень | Слабо умеет ставить новые задачи в области прикладной математики и информатики и находить пути их решения, а так же разрабатывать алгоритмы для написания компьютерных программ |
| Базовый уровень | Хорошо умеет ставить новые задачи в области прикладной математики и информатики и находить пути их решения, а так же разрабатывать алгоритмы для написания компьютерных программ |
| Профессиональный | Свободно умеет ставить новые задачи в области прикладной математики и информатики и находить пути их решения, а так же разрабатывать алгоритмы для написания компьютерных программ |
| ВЛАДЕТЬ:  различными методами, применяемыми при исследовании в области прикладной математики и информатики, в том числе владеть умением формулировать задачи, а так же разрабатывать алгоритмы и писать программы по данным алгоритмам | | Начальный уровень | Частично владеет различными методами, применяемыми при исследовании в области прикладной математики и информатики, в том числе формулирует задачи, а так же разрабатывает алгоритмы и пишет программы по данным алгоритмам |
| Базовый уровень | Достаточно хорошо владеет различными методами, применяемыми при исследовании в области прикладной математики и информатики, в том числе формулирует задачи, а так же разрабатывает алгоритмы и пишет программы по данным алгоритмам |
| Профессиональный | Свободно владеет различными методами, применяемыми при исследовании в области прикладной математики и информатики, в том числе формулирует задачи, а так же разрабатывает алгоритмы и пишет программы по данным алгоритмам |
| 5. | Характеристика взаимосвязи данной компетенции с другими компетенциями/ необходимость владения другими компетенциями для формирования данной компетенции | | Успешное формирование компетенции «способность анализировать и применять углубленные знания в области прикладной математики и информатики» базируется на освоении компетенций «способность к абстрактному мышлению, анализу, синтезу» и «способность самостоятельно приобретать с помощью информационных технологий и использовать в практической деятельности новые знания и умения, в том числе, в новых областях знаний, непосредственно не связанных со сферой деятельности, расширять и углублять свое научное мировоззрение» | |
| 6. | Средства и технологии оценки | | Набор практических заданий для выявления уровня сформированности компетенции | |

**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.) (при наличии)

-

**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

1. Филиал ОАО «РЖД»,

2. ОБУ «Информационно-технический центр»,

3. АО «НЛМК- Инжиниринг»,

4. Управление информатизации администрации Липецкой области,

5. ООО «МедСофт».

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

Развитие профессиональных качеств, сохранение и развитие квалификации.

Слушатель после окончания курса может выполнять следующие трудовые функции, связанные с анализом данных: Сбор и подготовка данных для анализа. Визуализация данных. Тестирование гипотез. Построение математических моделей процессов и систем и выбор адекватных им методов исследования. Анализ данных с применением языка R.

**VII.Дополнительная информация**

МФТИ (Машинное обучение и анализ данных) <https://mipt.ru/cdpo/programs/software/supervised_learning.php?ELEMENT_ID=1855780> 150 ч. 69600 руб.  
ВШЭ (Основы машинного обучения и продвинутые методы машинного обучения) <https://cs.hse.ru/dpo/advancedml> 114 ч. 88000 руб.  
Финансовый университет (Искусственный интеллект, большие данные и машинное обучение) <http://xn--80aaoligqe4brp0a.xn--p1ai/programms/iibd/> (краткосрочная программа) 19800 руб.

**VIII.Приложенные Скан-копии**

Утвержденной рабочей программа (подпись, печать, в формате pdf)