**Министерство образования и науки Республики Саха (Якутия)**

**Государственное автономное профессиональное образовательное учреждение Республики Саха (Якутия) «Якутский колледж связи и энергетики имени П.И. Дудкина»**

**ДОПОЛНИТЕЛЬНАЯ ПРОФЕССИОНАЛЬНАЯ ПРОГРАММА**

**ПОВЫШЕНИЯ КВАЛИФИКАЦИИ**

**«МАШИННОЕ ОБУЧЕНИЕ И БОЛЬШИЕ ДАННЫЕ»**

**72 ч**

**Форма обучения: очная и электронная с применением дистанционных технологий**

**2020 г.**

**ОБЩАЯ ХАРАКТЕРИСТИКА ПРОГРАММЫ**

**ДОПОЛНИТЕЛЬНАЯ ПРОФЕССИОНАЛЬНАЯ ПРОГРАММА**

**ПОВЫШЕНИЯ КВАЛИФИКАЦИИ**

по направлению «Информатика и вычислительная техника»

**«Машинное обучение и большие данные»**

1. **ОБЩАЯ ХАРАКТЕРИСТИКА ПРОГРАММЫ**

Дополнительная профессиональная программа (далее - ДПП) повышения квалификации разработана на основании:

- Федерального закона от 29.12.2012 г. № 27Э-ФЗ «Об образовании в Российской Федерации»,

- Приказа Министерства образования и науки Российской Федерации от 01.07. 2013 г. № 499 «Об утверждении Порядка организации и осуществления образовательной деятельности по дополнительным профессиональным программам»,

- Приказа Министерства образования и науки РФ от 23.08.2017 г. № 816 «Об утверждении Порядка применения организациями, осуществляющими образовательную деятельность, электронного обучения, дистанционных образовательных технологий при реализации образовательных программ»,

- Приказа Министерства образования и науки Российской Федерации от 02.07.2013 № 513 «Об утверждении Перечня профессий рабочих, должностей служащих, по которым осуществляется профессиональное обучение».

- Стандарт Ворлдскиллс (WSSS)по профессиональной компетенции «Машинное обучение и большие данные» (WorldSkills Standards Specifications). Техническое описание, спецификации стандарта.

- Федеральный государственный образовательный стандарт 09.02.07 «Информационные системы и программирование» от 09.12.2016 г. № 1547.

Трудоёмкость обучения: 72 академических часа.

Форма обучения: онлайн.

**1.Цель программы**

 Дополнительная профессиональная программа повышения квалификации направлена на совершенствование и (или) формирования у обучающегося новой компетенции в области цифровой экономики, необходимой для профессиональной деятельности.

**2.Планируемые результаты обучения:**

2.1.Знание (осведомленность в областях)

Основам Big Data и Data Science.

Основы языка программирования Python;

Основы анализа и визуализации данных, применяемый для этой цели базовый инструментарий языка Python;

Основные понятия и методики библиотеки numpy: массив, скаляр, массовая операция, основные операции с массивами, основные функции библиотеки;

Библиотеку Matplotlib ее основные понятия и методики применения;

Основные виды графиков и диаграмм, способы их построения;

Основные понятия и методики библиотеки Pandas.

2.2. Умение (способность к деятельности)

Основам работы с SQL — самым популярным языком структурированных запросов.

Как использовать язык программирования Python для анализа данных.

Как работать с данными, основы визуализации и описательного анализа.

Разовьете пространственное, статистическое и функциональное мышление.

Создавать и обрабатывать числовые массивы.

Создавать диаграммы и графики различных видов и форматов.

2.3.Навыки (использование конкретных инструментов)

Использование современных программных средств, предназначенных для машинного обучения и больших данных

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

* 1. Среднее профессиональное и (или) высшее профессиональное образование
	2. Квалификация
	3. Наличие опыта профессиональной деятельности
	4. Предварительное освоение иных дисциплин/курсов /модулей

**4.Учебный план программы «Машинное обучение и большие данные»**

|  |  |  |  |
| --- | --- | --- | --- |
| **№ п/п** | **Модуль**  | **Всего, час** | **Виды учебных занятий** |
| **лекции** | **практические занятия** | **самостоятельная работа** |
| 1 | Модуль 1. Введение в data science | **4** | 4 |  |  |
| 2 | Модуль 2. Основы SQL | **6** | 2 | 4 |  |
| 3 | Модуль 3. Python для анализа данных | **10** | 4 | 6 |  |
| 4 | Модуль 4. Работа с данными | **8** |  | 8 |  |
| 5 | Модуль 5. Визуализация | **4** |  | 4 |  |
| 6 | Модуль 6. Анализ данных | **12** | 4 | 8 |  |
| 7 | Модуль 7. Машинное обучение | **12** | 4 | 8 |  |
| 8 | Модуль 8. Подбор переменных и алгоритмов для модели | **12** | 4 | 8 |  |
| **Итоговая аттестация** |  **6** | **Зачёт** |

**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 | Модуль 1. Введение в data science | 4 |  |
| 2 | Модуль 2. Основы SQL | 6 |  |
| 3 | Модуль 3. Python для анализа данных | 10 |  |
| 4 | Модуль 4. Работа с данными | 8 |  |
| 5 | Модуль 5. Визуализация | 4 |  |
| 6 | Модуль 6. Анализ данных | 12 |  |
| 7 | Модуль 7. Машинное обучение | 12 |  |
| 8 | Модуль 8. Подбор переменных и алгоритмов для модели | 12 |  |
| 9 | **Итоговая аттестация** | 6 |  |
| **Всего:** | **72** | **От 2-х недель до 4-х недель** |

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

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **№ п/п** | **Модуль / Тема** | **Всего, час** | **Виды учебных занятий** | **Формы контроля** |
| **лекции** | **практические занятия** | **самостоятельная работа** |
| **1** | **Модуль 1. Введение в data science** | **4** | **4** |  |  | **Тест** |
| 1.1 | Введение в Big Data и Data science. | 2 | 2 |  |  |  |
| 1.2 | Математические приемы, используемые при работе с данными: матричные операции, введение в статистику и проверку гипотез. | 2 | 2 |  |  |  |
| **2** | **Модуль 2. Основы SQL** | **6** | **2** | **4** |  | **Тест****Выполненные практические работы** |
| 2.1 | SQL – язык структурированных запросов. | 2 | 2 |  |  |  |
| 2.2 | Базовая концепция, фильтрация, функции. | 2 |  | 2 |  |  |
| 2.3 | Подзапросы, Join, Табличные операции. | 2 |  | 2 |  |  |
| **3** | **Модуль 3. Python для анализа данных.** | **10** | **2** | **8** |  | **Тест****Выполненные практические работы** |
| 3.1 | Введение в язык программирования Python. | 1 | 1 |  |  |  |
| 3.2 | Основные функции. | 1 | 1 |  |  |  |
| 3.3 | Условные операторы. | 2 |  | 2 |  |  |
| 3.4 | Списки, словари, кортежи. | 2 |  | 2 |  |  |
| 3.5 | Циклы и условия. | 2 |  | 2 |  |  |
| 3.6 | Функции и классы. | 2 |  | 2 |  |  |
| **4** | **Модуль 4. Работа с данными** | **8** |  | **8** |  | **Выполненные практические работы** |
| 4.1 | Библиотека NumPy. | 2 |  | 2 |  |  |
| 4.2 | Матричные операции, семплирование, чтение файлов. | 2 |  | 2 |  |  |
| 4.3 | Библиотека SciPy. | 2 |  | 2 |  |  |
| 4.4 | Библиотека Pandas. | 2 |  | 2 |  |  |
| **5** | **Модуль 5. Визуализация** | **4** |  | **4** |  | **Выполненные практические работы** |
| 5.1 | Библиотека Matplotlib. |  |  | 1 |  |  |
| 5.2 | Библиотека Seaborn. |  |  | 1 |  |  |
| 5.3 | Библиотека Plotly. |  |  | 1 |  |  |
| 5.4 | Библиотека Dash. |  |  | 1 |  |  |
| **6** | **Модуль 6. Анализ данных** | **12** | **4** | **8** |  | **Выполненные практические работы** |
| 6.1 | Анализ данных | 3 | 1 | 2 |  |  |
| 6.2 | Визуализация | 3 | 1 | 2 |  |  |
| 6.3 | Гистрограммы распределения | 3 | 1 | 2 |  |  |
| 6.4 | Корреляционные матрицы | 3 | 1 | 2 |  |  |
| **7** | **Модуль 7. Машинное обучение** | **12** | **4** | **8** |  | **Выполненные практические работы** |
| 7.1 | Предобработка данных | 3 | 1 | 2 |  |  |
| 7.2 | Настройка гиперпараметров модели | 3 | 1 | 2 |  |  |
| 7.3 | Метод к-ближайших соседей | 3 | 1 | 2 |  |  |
| 7.4 | Регрессия | 3 | 1 | 2 |  |  |
| **8** | **Модуль 8. Подбор переменных и алгоритмов для модели** | **12** | **4** | **8** |  | **Выполненные практические работы** |
| 8.1 | Работа с категориальными признаками | 3 | 1 | 2 |  |  |
| 8.2 | Прогнозирование переменной | 3 | 1 | 2 |  |  |
| 8.3 | Оверсемплинг, Андерсемплинг | 3 | 1 | 2 |  |  |
| 8.4 | Lightgbm, Catboost | 3 | 1 | 2 |  |  |
| 9 | Итоговая аттестация | 6 |  |  | 6 | Итоговая работа |

**7. Учебная (рабочая) программа повышения квалификации «Машинное обучение и большие данные»**

Модуль 1. Введение в data science. (4 час.)

Тема 1.1 Введение в Big Data и Data science. (2 час.)

Тема 1.2 Математические приемы, используемые при работе с данными: матричные операции, введение в статистику и проверку гипотез. (2 час.)

Модуль 2. Основы SQL. (6 час.)

Тема 2.1 SQL – язык структурированных запросов. (2 час.)

Тема 2.2 Базовая концепция, фильтрация, функции. (2 час.)

Тема 2.3 Подзапросы, Join, Табличные операции. (2 час.)

Модуль 3. Python для анализа данных. (10 час.)

Тема 3.1 Введение в язык программирования Python. (1 час.)

Тема 3.2 Основные функции. (1 час.)

Тема 3.3 Условные операторы. (2 час.)

Тема 3.4 Списки, словари, кортежи. (2 час.)

Тема 3.5 Циклы и условия. (2 час.)

Тема 3.6 Функции и классы. (2 час.)

Модуль 4. Работа с данными (8 час.)

Тема 4.1 Библиотека NumPy. (2 час.)

Тема 4.2 Матричные операции, семплирование, чтение файлов. (2 час.)

Тема 4.3 Библиотека SciPy. (2 час.)

Тема 4.4 Библиотека Pandas. (2 час.)

Модуль 5. Визуализация (4 час.)

Тема 5.1 Библиотека Matplotlib. (1 час.)

Тема 5.2 Библиотека Seaborn. (1 час.)

Тема 5.3 Библиотека Plotly. (1 час.)

Тема 5.4 Библиотека Dash. (1 час.)

Модуль 6. Анализ данных. (12 час.)

Тема 6.1 Анализ данных. (3 час.)

Тема 6.2 Визуализация. (3 час.)

Тема 6.3 Гистрограммы распределения. (3 час.)

Тема 6.4 Корреляционные матрицы. (3 час.)

Модуль 7. Машинное обучение. (12 час.)

Тема 7.1 Предобработка данных. (3 час.)

Тема 7.2 Настройка гиперпараметров модели. (3 час.)

Тема 7.3 Метод к-ближайших соседей. (3 час.)

Тема 7.4 Регрессия. (3 час.)

Модуль 8. Подбор переменных и алгоритмов для модели. (12 час.)

Тема 8.1 Работа с категориальными признаками. (3 час.)

Тема 8.2 Прогнозирование переменной. (3 час.)

Тема 8.3 Оверсемплинг, Андерсемплинг. (3 час.)

Тема 8.4 Lightgbm, Catboost. (3 час.)

**Описание практико-ориентированных заданий и кейсов**

|  |  |  |  |
| --- | --- | --- | --- |
| **№** | **Номер темы/модуля** | **Наименование практического занятия** | **Описание** |
| 2 | Модуль 2. Основы SQL |  |  |
|  | 2.2 | Базовая концепция, фильтрация, функции. | Создание фильтров |
|  | 2.3 | Подзапросы, Join, Табличные операции. | Создание запросов, табличных операций |
| 3 | Модуль 3. Python для анализа данных. |  |  |
|  | 3.3 | Условные операторы. | Решение задач с использованием условного оператора |
|  | 3.4 | Списки, словари, кортежи. | Решение задач на тему списки, словари, кортежи |
|  | 3.5 | Циклы и условия. | Решение задач с использованием циклов и условий |
|  | 3.6 | Функции и классы. | Решение задач с использованием функции и классов |
| 4 | Модуль 4. Работа с данными |  |  |
|  | 4.1 | Библиотека NumPy. | Типы данных, массивы и операции с ними. |
|  | 4.2 | Матричные операции, сэмплирование, чтение файлов. | Работа с файлами |
|  | 4.3 | Библиотека SciPy.  | Научные вычисления. Работа с данными. Тестирование данных. |
|  | 4.4 | Библиотека Pandas. | Загрузка и запись данных, срезы данных, мультииндексация, groupby, datetime, статистические функции. |
| 5 | Модуль 5. Визуализация |  |  |
|  | 5.1 | Библиотека Matplotlib | Варианты отрисовки графиков и изображений. |
|  | 5.2 | Библиотека Seaborn,  | Визуализация парных взаимосвязей, heatmap, диаграммы. |
|  | 5.4 | Библиотека Plotly.  | Продвинутая визуализация. Динамические графики. |
|  | 5.5 | Библиотека Dash.  | Обзор фреймворка для создания дэшбордов (аналитические BI инструменты). |
| 6 | Модуль 6. Загрузка, подготовка и предварительное изучение данных |  |  |
|  | 6.1 | Анализ данных | Решение практических задач по анализам данных |
|  | 6.2 | Визуализация | Решение практических задач по построению графиков одномерных и многомерных функций распределения, точечных диаграмм, тепловых карт |
|  | 6.3 | Гистограммы распределения | Решение практических задач по построению гистограммы распределения |
|  | 6.4 | Корреляционные матрицы | Решение практических задач по корреляционным матрицам |
| 7 | Модуль 7. Изучение параметров, влияющих на целевую переменную. Предварительная подготовка данных |  |  |
|  | 7.1 | Предобработка данных | Предобработка данных. Масштабирование. Разбиение данных на выборки |
|  | 7.2 | Настройка гиперпараметров модели | Подбор признаков и оптимизация гиперпараметров |
|  | 7.3 | Метод к-ближайших соседей | Описание метода kNN.Область решаемых задач.Метрики качества.Разбор кода с примером реализации алгоритма. |
|  | 7.4 | Регрессия | Задача регрессии.Линейная регрессия, логистическая регрессия.Регрессии: разбор кода с реализацией алгоритма. |
| 8 | Модуль 8. Подбор переменных и алгоритмов для модели |  |  |
|  | 8.1 | Работа с категориальными признаками | Закодирование переменных с помощью словаря Label EncodingOne-Hot EncodingХэширование признаков (Hashing trick) |
|  | 8.2 | Прогнозирование переменной | Разбиение данных.Прогнозирование переменной |
|  | 8.3 | Оверсемплинг, Андерсемплинг | Оверсемплинг и андерсемплинг в задачах классификации |
|  | 8.4 | Lightgbm, Catboost | Lightgbm, Catboost |

**8.Оценочные материалы по образовательной программе**

**8.1. Тесты в конце модуля**

**8.2 Практико-ориентированные задания и кейсы**

**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** | Иванова Сардана Кимовна | ГАПОУ РС(Я) Якутский колледж связи и энергетики им.П.И.Дудкина |  | **+** | **+** |

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

|  |
| --- |
| **Учебно-методические материалы** |
| Методы, формы и технологии | Методические разработки, материалы курса, учебная литература |
| Электронное обучение с применением дистанционных технологий | 1. Уэс Маккинли. Python и анализ данных;
2. Эрик Мэтиз. Изучаем Python
3. Прохоренок Н.А. Python 3. Самое необходимое
4. Бринк Хенрик, Ричард Джозеф, Феверолф Марк. Машинное обучение
5. Бенгфорт Бенджамин, Билбро Ребекка, Охеда Тони. Прикладной анализ текстовых данных на Python. Машинное обучение и создание приложений обработки естественного языка
6. Плас Дж. Вандер. Python для сложных задач: наука о данных и машинное обучение.
 |

|  |
| --- |
| **Информационное сопровождение** |
| Электронные образовательные ресурсы | Электронные информационные ресурсы |
| do.yakse.ru |  |
|  |  |

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

|  |  |
| --- | --- |
| Вид занятий | Наименование оборудования, программного обеспечения |
| Практические работы в мастерской по компетенции «Машинное обучение и большие данные» соответствующей международным стандартам Ворлдскиллс | Оборудование по спецификации компетенции |

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

|  |  |
| --- | --- |
| Вид занятий | Наименование оборудования, программного обеспечения |
| Практические работы в мастерской по компетенции «Машинное обучение и большие данные» соответствующей международным стандартам Ворлдскиллс | Оборудование по спецификации компетенции |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| № п/п | Наименование ТМЦ | Кол-во | Дата установки/монтажа | Инвентарный № | Техническое состояние |
| 1 | Стол ученический 2100х850х750 | 9 | 23.09.2020 |  | новое |
| 2 | Стол преподавателя 1600х850х750 | 1 | 23.09.2020 |  | новое |
| 3 | Кресло «Престиж» кож/зам синего цвета без ручек | 18 | 23.09.2020 |  | новое |
| 4 | Кресло «Престиж» кож/зам синего цвета с рeчrами | 1 | 23.09.2020 |  | новое |
| 5 | Шкаф металлический «Архив» | 1 | 23.09.2020 |  | новое |
| 6 | Шкаф ШРМ-312  | 1 | 23.09.2020 |  | новое |
| 7 | **Компьютер в комплекте** Intel Core i9 9‑го поколения с тактовой частотой 4,0 ГГц (ускорение Turbo Boost до 4,1 ГГц) 64 ГБ памяти DDR4 2400 МГц Графический процессор GEforce RTX 2060 Super с 8 ГБ памяти GDDR5, SSD 490 Gb, HDD 3TbМонитор 24" по 2 монитора Клавиатура и мышьинтерфейсный кабель HDMI | 19 | 23.09.2020 |  | новое |
| 8 | 15.6" Ноутбук Dell Triton 300 PT315-51-77K2 черный | 1 |  |  |  |
| 9 | Проектор Epson EH-TW650 | 1 | 23.09.2020 |  | новое |
| 10 | Экран для проектора Sakura 300x300 Motoscreen 1:1 167'' фибергласс, Gray (SCPSM-300X300FG-GR) | 1 | 23.09.2020 |  | новое |
| 13 | DS-l252 Купольная IP-видеокамера с ИК-подсветкой до 30м | 1 | 23.09.2020 |  | новое |
| 14 | DS-l214 Компактная IP-видеокамера с ИК-подсветкой до 10м | 1 | 23.09.2020 |  | новое |
| 15 | Коммутатор 16 портов ELTEX | 1 |  |  |  |
| 16 | Коммутатор 16 портов TPLink | 1 |  |  |  |

Программное обеспечение

|  |  |  |
| --- | --- | --- |
| № | наименование | количество |
| 1 | RStudio | 25 |
| 2 | ПО Офисный пакет (Word, Excel, PowerPoint) 2019 | - |
| 3 | Windows 10 pro | - |

Имеется доступ к Серверам колледжа (1 Gb), есть доступ к сети Интернет.

Перечень составлен согласно инфраструктурному листу отборочных соревнований Молодые профессионалы (Ворлдскиллс Россия).

**Литература**

1. Златопольский Д.М. Основы программирования на языке Python. – М.: ДМК Пресс, 2017. – 284 с.
2. Гэддис Т. Начинаем программировать на Python.  – 4-е изд.: Пер. с англ. – СПб.: БХВ-Петербург, 2019. – 768 с.
3. Лучано Рамальо Python. К вершинам мастерства. – М.: ДМК Пресс, 2016. – 768 с.
4. Свейгарт, Эл. Автоматизация рутиных задач с помощью Python: практическое руководство для начинающих. Пер. с англ. — М.: Вильямc, 2016. – 592 с.
5. Рейтц К., Шлюссер Т. Автостопом по Python. – СПб.: Питер, 2017. – 336 с.: ил. – (Серия «Бестселлеры O’Reilly»).
6. Любанович Билл Простой Python. Современный стиль программирования. – СПб.: Питер, 2016. – 480 с.: – (Серия «Бестсепперы O’Reilly»).
7. Федоров, Д. Ю.  Программирование на языке высокого уровня Python : учебное пособие для прикладного бакалавриата / Д. Ю. Федоров. – 2-е изд., перераб. и доп. – Москва : Издательство Юрайт, 2019. – 161 с. – (Бакалавр. Прикладной курс). – ISBN 978-5-534-10971-9. – Текст: электронный // ЭБС Юра

Составитель: Иванова С.К.