

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ
РОССИЙСКОЙ ФЕДЕРАЦИИ
ФГАОУ ВО «СИБИРСКИЙ ФЕДЕРАЛЬНЫЙ УНИВЕРСИТЕТ»



УТВЕРЖДАЮ

Директор НОЦ «Институт
непрерывного образования»

Е.В. Мошкина Е.В. Мошкина

марта 2024 г.

ДОПОЛНИТЕЛЬНАЯ ПРОФЕССИОНАЛЬНАЯ ПРОГРАММА
ПОВЫШЕНИЯ КВАЛИФИКАЦИИ

«Анализ данных в промышленности»

Красноярск 2024

I. ОБЩАЯ ХАРАКТЕРИСТИКА ПРОГРАММЫ

1.1. Аннотация программы

Программа повышения квалификации «Анализ данных в промышленности» предназначена для слушателей, которые хотят освоить основы анализа данных и применить их на практике. Слушатели изучат математические концепции, лежащие в основе алгоритмов машинного обучения и базовые алгоритмы машинного обучения для анализа данных в промышленности и принятия обоснованных решений.

Практическая часть программы направлена на разбор различных кейсов в области анализа данных.

1.2. Цель программы

Цель программы повышения квалификации — формирование и(или) совершенствование компетенций, необходимых для выполнения трудовых функций и профессиональной деятельности в области машинного обучения и анализа данных в соответствии с требованиями профессионального стандарта 06.042 «Специалист по большим данным».

1.3. Компетенции (трудовые функции) в соответствии с Профессиональным стандартом (формирование новых или совершенствование имеющихся)

Программа ориентирована на формирование компетенций слушателей в области машинного обучения и анализа данных по обобщенным трудовым функциям профессионального стандарта 06.042 «Специалист по большим данным» (утвержден приказом Министерства труда и социальной защиты Российской Федерации от 6 июля 2020 года N 405н):

А. Анализ больших данных с использованием существующей в организации методологической и технологической инфраструктуры:

А/02.6 Планирование и организация аналитических работ с использованием технологий больших данных.

А/03.6 Подготовка данных для проведения аналитических работ по исследованию больших данных.

А/04.6 Проведение аналитического исследования с применением технологий больших данных в соответствии с требованиями заказчика.

1.4. Планируемые результаты обучения

В результате освоения программы слушатели будут обладать профессиональными компетенциями, включающими в себя способность:

РО1. Понимать математические основы алгоритмов машинного обучения и методов анализа данных и вычислять необходимые статистики (А/04.6).

РО2. Применять для решения задач анализа данных средства таких библиотек, как NumPy, SciPy, Matplotlib, Pandas и Scikit-learn (А/03.6).

РО3. Формулировать и решать задачи обучения на размеченных данных с помощью различных методов машинного обучения (А/02.6).

РО4. Оценивать качество и адекватность моделей машинного обучения применительно к решаемым задачам (А/04.6).

РО5. Реализовывать полный цикл работы с данными, включая их сбор, подготовку, описательный и визуальный анализ, подбор и реализацию подходящих методов анализа, оценку и интерпретацию результатов, презентацию итогов проекта (А/03.6, А/04.6).

1.5. Категория слушателей

Студенты, обучающиеся по программам бакалавриата и специалитета в области ИТ и промышленности.

1.6. Требования к уровню подготовки поступающего на обучение

В соответствии с требованиями к образованию и обучению, предъявляемыми к 4 уровню квалификации профессионального стандарта 06.042 «Специалист по большим данным», необходимо иметь высшее образование (бакалавриат).

Базовые знания математики и основ программирования.

1.7. Продолжительность обучения

Повышение квалификации – 108 часов.

1.8. Форма обучения

Очно-заочная с использованием электронного обучения и дистанционных образовательных технологий.

1.9. Требования к материально-техническому обеспечению, необходимому для реализации дополнительной профессиональной программы повышения квалификации (требования к аудитории, компьютерному классу, программному обеспечению)

Компьютерный мультимедийный класс на 15-20 ПК со стабильным интернет-соединением, в том числе через wi-fi; маркерная доска; проектор с возможностью подключения ноутбука; компьютеры с установленным ПО для студентов: Python 3.

1.10. Документ об образовании: удостоверение о повышении квалификации установленного образца.

УЧЕБНО-ТЕМАТИЧЕСКИЙ ПЛАН
программы повышения квалификации
«Анализ данных в промышленности»

Категория слушателей: лица, имеющие/получающие высшее образование.

Срок обучения: 12 недель.

Форма обучения: очно-заочная с использованием электронного обучения и дистанционных образовательных технологий.

Режим занятий: 2–3 часа в день.

№ п/п	Наименование и содержание разделов и тем программы	Общая трудоемкость, ч	Всего контактн., ч	Контактные часы		СРС, ч.	Использование средств ЭО и ДОТ	Результаты обучения
				Лекции	Практ. и семинарские занятия			
1.	Математика машинного обучения	36	18	-	18	18		PO1, PO2
1.1.	Основы линейной алгебры	8	4	-	4	4	Материалы электронного курса: тест, файл, задание	PO1, PO2
1.2.	Методы оптимизации	8	4	-	4	4		PO1, PO2
1.3.	Теория вероятностей	8	4	-	4	4		PO1, PO2
1.3.1	Событие. Вероятность события	4	2	-	2	2		PO1, PO2
1.3.2	Случайные величины. Числовые характеристики случайной величины. Распределения случайных величин	4	2	-	2	2		PO1, PO2
1.4.	Математическая статистика	12	6	-	6	6		PO1, PO2
1.4.1	Важные статистики	4	2	-	2	2		PO1, PO2
1.4.2	Оценка распределения по выборке	4	2	-	2	2		PO1, PO2
1.4.3	Доверительные интервалы	4	2	-	2	2		PO1, PO2
2.	Базовые алгоритмы машинного обучения и их применение	68	34	-	34	34		
2.1	Python для анализа данных	8	4	-	4	4		PO2, PO5
2.1.1	Введение в анализ данных. Знакомство с Python	4	2	-	2	2		PO2, PO5

№ п/п	Наименование и содержание разделов и тем программы	Общая трудоемкость, ч	Всего контактн., ч	Контактные часы		СРС, ч.	Использование средств ЭО и ДОТ	Результаты обучения
				Лекции	Практ. и семинарские занятия			
2.1.2	Предварительная обработка и разведочный анализ данных	4	2	-	2	2	Материалы электронного курса: тест, файл, задание	PO2, PO5
2.2	Основы машинного обучения	48	24	-	24	24		PO2–PO5
2.2.1	Обучение на размеченных данных. Задача регрессии	8	4	-	4	4		PO2–PO4
2.2.2	Задача классификации. Линейные модели	4	2	-	2	2		PO2–PO4
2.2.3	Проблемы линейных моделей	4	2	-	2	2		PO2–PO4
2.2.4	Метод k ближайших соседей	4	2	-	2	2		PO2–PO4
2.2.5	Решающие деревья и композиции алгоритмов	4	2	-	2	2		PO2–PO4
2.2.6	Нейронные сети	8	4	-	4	4		PO2–PO4
2.2.7	Задача кластеризации	4	2	-	2	2		PO2, PO4
2.2.8	Иерархическая кластеризация	4	2	-	2	2		PO2, PO4
2.2.9	Метод главных компонент	4	2	-	2	2		PO2, PO4
2.2.10	Анализ временных рядов	4	2	-	2	2		PO2, PO4
2.3	Практические кейсы анализа данных	12	6	-	6	6		PO2, PO5
	Итоговая аттестация	4	2	-	2	2	PO1–PO5	
	ИТОГО	108	54	-	54	54		

II. ОСНОВНОЕ СОДЕРЖАНИЕ ПРОГРАММЫ

2.1. План учебной деятельности

Результаты обучения	Учебные действия/ формы текущего контроля	Используемые ресурсы/ инструменты/технологии
PO1. Понимать математические основы алгоритмов машинного обучения и методов анализа данных и вычислять необходимые статистики	<i>На семинарских занятиях:</i> Составление конспектов, решение практических задач, дискуссии. <i>В онлайн-курсе:</i> Изучение теоретического материала, выполнение тренировочных тестов, выполнение практических заданий	Материалы электронного курса в системе электронного обучения СФУ «e-Курсы»
PO2. Применять для решения задач анализа данных средства таких библиотек, как NumPy, SciPy, Matplotlib, Pandas и Scikit-learn	<i>На семинарских занятиях:</i> Решение практических задач, разбор кейсов. <i>В онлайн-курсе:</i> Изучение теоретического материала, выполнение тренировочных тестов, выполнение практических заданий, выполнение заданий на программирование	Материалы электронного курса в системе электронного обучения СФУ «e-Курсы». Jupyter notebook, Google colab
PO3. Формулировать и решать задачи обучения на размеченных данных с помощью различных методов машинного обучения	<i>На семинарских занятиях:</i> Составление конспектов, разбор кейсов, решение практических задач. <i>В онлайн-курсе:</i> Изучение теоретического материала, выполнение тренировочных тестов, выполнение заданий на программирование	Материалы электронного курса в системе электронного обучения СФУ «e-Курсы». Jupyter notebook, Google colab
PO4. Оценивать качество и адекватность моделей машинного обучения применительно к решаемым задачам	<i>На семинарских занятиях:</i> Составление конспектов, разбор кейсов, решение практических задач. <i>В онлайн-курсе:</i> Изучение теоретического материала, выполнение тренировочных тестов, выполнение заданий на программирование	Материалы электронного курса в системе электронного обучения СФУ «e-Курсы». Jupyter notebook, Google colab
PO5. Реализовывать полный цикл работы с данными, включая их сбор, подготовку, описательный и визуальный анализ, подбор и реализацию подходящих методов анализа, оценку и интерпретацию результатов, презентацию итогов проекта	<i>На семинарских занятиях:</i> Разбор кейсов, выполнение проекта, дискуссии. <i>В онлайн-курсе:</i> Выполнение заданий на программирование	Материалы электронного курса в системе электронного обучения СФУ «e-Курсы». Jupyter notebook, Google colab

2.2. Виды и содержание самостоятельной работы

Выполнение самостоятельной работы слушателями предполагается в дистанционном режиме в рамках электронных курсов «Математика машинного обучения» (<https://e.sfu-kras.ru/course/view.php?id=37735>) и «Базовые алгоритмы машинного обучения» (<https://e.sfu-kras.ru/course/view.php?id=37678>), размещенных в системе электронного обучения СФУ.

Самостоятельно слушателями изучаются представленный теоретический материал, кейсы и задачи, примеры по машинному обучению и анализу данных, выполняются задания по программированию, и тренировочные тесты.

III. УЧЕБНО-МЕТОДИЧЕСКИЕ МАТЕРИАЛЫ

3.1. Учебно-методическое обеспечение, в т.ч. электронные ресурсы в корпоративной сети СФУ и сети Интернет

Основная литература

1. Гасников А. В. Современные численные методы оптимизации. Метод универсального градиентного спуска: учебное пособие. – М.: МФТИ, 2018.

2. Гмурман В. Е. Теория вероятностей и математическая статистика: учеб. пособие для вузов. – М.: Юрайт, 2010.

3. Груздев А. В. Предварительная подготовка данных в Python. Т. 2. План, примеры и метрики качества. – М.: ДМК Пресс, 2023. – 814 с.

4. Груздев А.В. Предварительная подготовка данных в Python. Т. 1. Инструменты и валидация. – М.: ДМК Пресс, 2023. – 816 с.

5. Кудрявцев Л.Д. Краткий курс математического анализа. Т. 1–2. – М.: Физматлит, 2015.

6. Пантелеев А.В., Летова Т.А. Методы оптимизации в примерах и задачах: учебное пособие. – СПб.: Лань, 2015.

7. Плас Дж. Вандер Python для сложных задач: наука о данных и машинное обучение. – СПб.: Питер, 2018. – 576 с.

8. Рашка С. Python и машинное обучение: крайне необходимое пособие по новейшей предсказательной аналитике, обязательное для более глубокого понимания методологии машинного обучения. – М.: ДМК Пресс, 2017.

9. Хасты Т., Тибширани Р., Фридман Дж. Основы статистического обучения: интеллектуальный анализ данных, логический вывод и прогнозирование. – СПб.: ООО «Диалектика», 2020. – 768 с.

Дополнительная литература

1. Вентцель Е. С. Теория вероятностей: учебник для вузов. – М.: КноРус, 2010.

2. Златопольский Д. М. Основы программирования на языке Python. – 2-е изд. – М.: ДМК Пресс, 2018. – 396 с.

3. Курош А. Г. Курс высшей алгебры: учебник для вузов. – СПб.: Лань, 2008.

4. Маккинни, Уэс. Python и анализ данных / У. Маккинни. – М.: ООО «ЛитРес». – 484 с.

5. Рамальо, Лучано. Python. К вершинам мастерства / Л. Рамальо. – М.: ООО «ЛитРес». – 770 с.

3.2. Программное обеспечение (информационные обучающие системы, системы вебинаров, сетевые ресурсы хостинга видео, изображений, файлов, презентаций и др.)

1. Сервис видеоконференций Jazz.
2. Интернет-браузер.
3. Anaconda Individual Edition (with Python 3.7 и выше) с jupyter notebook.
4. Google colab: <https://colab.research.google.com/>.

IV. ОЦЕНКА КАЧЕСТВА ОСВОЕНИЯ ПРОГРАММЫ

4.1. Формы аттестации, оценочные материалы, методические материалы

Программа предусматривает проведение текущей и итоговой аттестации. Текущая аттестация слушателей проводится на основе оценки активности и участия в дискуссиях в ходе семинаров, а также качества выполнения заданий и тестов в электронных обучающих курсах.

Обучение на программе повышения квалификации предполагает выполнение индивидуальных текущих заданий и тестирований. Методические материалы, необходимые для выполнения текущих заданий, представлены в соответствующих элементах электронных обучающих курсов и включают описание задания, методические рекомендации по его выполнению, критерии оценивания.

Промежуточная аттестация проводится в формате тестирования. Основанием для прохождения промежуточной аттестации является получение не менее 60 % баллов за тестирование по модулям программы.

Реализация программы повышения квалификации завершается итоговой аттестацией в форме зачета. Ответ на зачете оценивается по бальной шкале. Итоговая оценка выставляется в соответствии с количеством баллов, набранных в ходе зачета.

4.2. Требования и содержание итоговой аттестации

К итоговой аттестации допускаются слушатели, выполнившие учебный план программы.

Итоговый зачет включает материалы всех изученных модулей и разделов. Содержание вопросов в билете требует умения формулировать ответы в свободной форме, работать с данными, ориентироваться в ключевых определениях и терминах.

Перечень вопросов для подготовки к итоговой аттестации

1. Какие матричные операции используются в линейной алгебре для работы с данными?
2. Что такое градиентный спуск и как он применяется в машинном обучении?
3. Какие основные законы вероятности и статистики необходимо знать для анализа данных?
4. Какие библиотеки Python чаще всего используются для работы с данными?
5. Чем отличается обучение с учителем от обучения без учителя?
6. Какие методы обучения без учителя вы знаете и в каких случаях их применяют?
7. В чем заключается принцип работы нейронных сетей?

8. Как можно анализировать временные ряды и выявлять в них закономерности?

9. Какие метрики качества моделей используются при обучении с учителем?

10. Какие методы предобработки данных используются перед построением модели машинного обучения?

11. Какие методы уменьшения размерности данных вы знаете?

12. Что такое кросс-валидация и как она помогает оценить качество модели?

13. Какие функции активации чаще всего используются в нейронных сетях?

14. Что такое кластеризация и как она применяется в анализе данных?

15. Как можно провести визуализацию данных для лучшего понимания их структуры?

16. Какие методы оптимизации используются при обучении нейронных сетей?

17. Какие методы ансамблевого обучения вы знаете и в чем их особенности?

18. Какие методы заполнения пропущенных данных вы знаете и как выбрать оптимальный?

19. Что такое регрессия и какие модели регрессии применяются в анализе данных?

20. Что такое переобучение и как его можно предотвратить?

21. Как можно выбрать оптимальное количество кластеров в задаче кластеризации?

22. Какие методы обработки выбросов данных вы знаете?

23. Что такое функция потерь и как она влияет на обучение модели?

24. Как проводится разведочный анализ данных и зачем он нужен?

25. Какие методы оценки значимости признаков существуют и какой из них лучше использовать?

26. Что представляют из себя свёрточные нейронные сети и в каких областях они применяются?

27. Какие методы предсказания временных рядов существуют и как выбрать подходящий?

28. Что такое регуляризация и зачем она применяется в машинном обучении?

29. Какие элементарные операции над матрицами вы знаете и как они используются в линейной алгебре?

30. Как можно решить систему линейных уравнений с помощью методов линейной алгебры?

31. Как можно оценить параметры распределения вероятности на основе выборки данных?

32. Какие методы обучения без учителя применяются для кластеризации данных?

33. Что такое глубокое обучение и в чем отличие от классических методов машинного обучения?

34. Как можно выявить сезонность и тренды в анализе временных рядов?

35. Что такое мультиколлинеарность и как она влияет на модели регрессии?

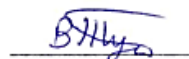
Программу составили:

Канд. физ.-мат. наук, доцент кафедры
прикладной математики и анализа данных




А.С. Кацунова

Старший преподаватель кафедры
прикладной математики и анализа данных



В.С. Тутатчиков

Канд. пед. наук, доцент кафедры
прикладной математики и анализа данных



Р.В. Есин

Руководитель программы:

Канд. пед. наук, доцент кафедры
прикладной математики и анализа данных



Р.В. Есин