Общество с ограниченной ответственностью «ЮМАКС»

УТВЕРЖДЕНО

Генеральный директор ООО «Юмакс»

_____/М. Ю. Мягков/

01 сентября 2024 года



Дополнительная профессиональная программа повышения квалификации

«Ведение в анализ данных и машинное обучение»

(80 часов)

Автор(ы):
Булгакова Т. Н., методист-эксперт

Москва, 2024

Раздел 1. «Характеристика программы»

1.1. Цель реализации программы

Совершенствование имеющихся у слушателей компетенций и формирование новых компетенций в области искусственного интеллекта, анализа данных и машинного обучения.

1.2. Характеристика совершенствуемых и/или получаемых новых компетенций, необходимых для профессиональной деятельности и/или повышения профессионального уровня в рамках имеющейся квалификации

Компетенции, формируемые и совершенствуемые в результате обучения, представлены в таблицах 1 и 2.

Таблица 1

№ п/п	Компетенции в соответствии с направлением подготовки 09.03.02 «Информационные системы и технологии» (уровень бакалавриата)	Код компетенции из ФГОС
1.	Способен использовать современные информационные технологии и программные средства, в том числе отечественного производства, при решении задач профессиональной деятельности;	ОПК-2
2.	Способен разрабатывать алгоритмы и программы, пригодные для практического применения в области информационных систем и технологий;	ОПК-6

Таблица 2

№ п/п	Компетенция в соответствии с профессиональным стандартом 06.042 «Специалист по большим данным»	Код компетенции
1.	Способен выполнять подготовку данных для проведения аналитических работ по исследованию больших данных;	A/03.6
2.	Способен выполнять проведение аналитического исследования с применением технологий больших данных в соответствии с требованиями заказчика;	A/04.6

1.3. Описание трудовых функций, входящих в профессиональный стандарт (функциональная карта вида профессиональной деятельности)

Программа разработана в соответствии с профессиональным стандартом

«Специалист по большим данным».

Обобщенные трудовые функции			Трудовые функции		
код	**наименование**	**уровень квалификации**	**наименование**	**код**	**уровень (подуровень) квалификации**
A	Анализ больших данных с использованием существующей в организации методологической и технологической инфраструктуры	6	Подготовка данных для проведения аналитических работ по исследованию больших данных.	A/03.6	6
		6	Проведение аналитического исследования с применением технологий больших данных в соответствии с требованиями заказчика	A/04.6	6

1.4. Планируемые результаты обучения (теоретические и профессиональные)

1.4.1. Теоретические

№ п/п	**Содержание компетенции**	**09.03.02 «Информационные системы и технологии» (уровень бакалавриата)**
		Код компетенции
	Знать: - Основные этапы работы с данными - Актуальные методы и инструменты для анализа данных - Технологические решения и программное обеспечение для анализа данных *Уметь:* - Реализовывать аналитический процесс от данных до интерпретации результата	ОПК-2

	- Использовать методы и инструменты анализа данных для решения практических задач *Владеть:* - Инструментами для анализа данных и представления результатов в наглядной форме	
	Знать: - Принципы организации работы с проектами в области машинного обучения *Уметь:* - Применять принципы воспроизводимости экспериментов и отчетности *Владеть:* - Инструментальными средствами разработки алгоритмов машинного обучения - Принципами командной и проектной работы в области анализа данных и машинного обучения	ОПК-6
		Профессиональный стандарт 06.042 «Специалист по большим данным»
		Код компетенции
	Знать: - Методы работы с различными структурами данных - Библиотеки Python для анализа данных (NumPy, Pandas, Matplotlib/Seaborn) - Feature Engineering: основные методы и подходы генерации и отбора признаков - Feature Engineering: принципы кодирования категориальных признаков и работы с разреженными данными *Уметь:* - Осуществлять предварительный анализ данных, выявлять закономерности и аномалии - Готовить данные для дальнейшего использования (обработка пропусков, масштабирование, нормализация) - Feature Engineering: проводить предобработку признаков (Label Encoding, One-hot Encoding и др.) - Feature Engineering: реализовывать методы выбора признаков, учитывая целевую переменную	A/03.6

	Владеть: - Навыками разведочного анализа данных (EDA) - Приемами визуализации данных для выявления зависимостей - Feature Engineering: навыками построения и тестирования новых признаков для повышения производительности модели	
	Знать: - Основные алгоритмы машинного обучения (линейные модели, деревья решений, ансамблевые методы) - Методы обучения с учителем и без учителя - Как гиперпараметры влияют на производительность алгоритмов - Основные методы оптимизации гиперпараметров (GridSearch, RandomSearch) - Полный жизненный цикл разработки моделей машинного обучения - Ключевые этапы разработки моделей от анализа данных до интерпретации результатов *Уметь*: - Реализовывать алгоритмы классификации и регрессии с использованием Scikit-learn - Применять ансамблевые подходы (Random Forest, Gradient Boosting) - Настраивать гиперпараметры с использованием оптимизированных поисковых подходов - Работать с методами регуляризации для улучшения обобщающих свойств моделей - Создавать и тестировать модели для реальных практических задач - Разрабатывать сквозные решения для обработки данных и построения метрик *Владеть:* - Подходами к выбору моделей в зависимости от задач и данных - Навыками оценки качества произведенных моделей (метрики) - Инструментами для автоматизации поиска и фиксации конфигураций гиперпараметров - Навыками Python-программирования и его приложений в задачах машинного обучения	A/04.6

1.4.2. Профессиональные

1.4.2.1 Трудовые действия (в соответствии с компетенциями A/03.6, A/04.6):

1. Извлечение, проверка и очистка больших объемов данных из гетерогенных источников
2. Агрегация и разработка представления больших объемов данных из гетерогенных источников
3. Оценка соответствия набора данных предметной области и задачам аналитических работ
4. Выбор методов и инструментальных средств анализа больших данных для проведения аналитических работ.
5. Разработка, поверка, оценка используемых моделей больших данных.
6. Выбор средств представления результатов аналитики больших данных.

1.5. Категория слушателей.

Программа повышения квалификации предназначена для специалистов, имеющих высшее или среднее профессиональное образование, а также лиц, получающих высшее или среднее профессиональное образование. К обучению на программе допускаются: работники компаний IT-сектора, а также мотивированные специалисты из других профессиональных сфер и студенты, обучающиеся в области информационных технологий, а также по иным специальностям, которые заинтересованы в получении новых компетенций по специальности «Аналитик данных и машинное обучение»

Область профессиональной деятельности: Связь, информационные и коммуникационные технологии

1.6. Форма обучения: очная форма с применением дистанционных образовательных технологий

1.7. Режим занятий: онлайн

1.8. Трудоемкость программы: 80 ак.ч.

Раздел 2. «Содержание программы»

2.1. Календарный учебный график

месяцы	1 месяц				2 месяц				3 месяц				4 месяц			
недели	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Лекции	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
Практические занятия	4	4	2	4	2	2	4	2	4	4	3	3	4	2	2	2

2.2. Учебный план

№ п/п	Тема	Всего часов	В том числе		Форма контроля
			Лекции	Практические занятия	
1	Введение в анализ данных и машинное обучение	12	4	8	Экспертная оценка преподавателя, проверочные тесты
2	Работа с данными	14	6	8	Экспертная оценка преподавателя
3	Основы машинного обучения	14	6	8	Экспертная оценка преподавателя, проверочные тесты
4	Продвинутые методы машинного обучения	12	4	8	Экспертная оценка преподавателя
5	Введение в нейронные сети	16	6	10	Экспертная оценка преподавателя, проверочные тесты
6	Итоговый проект	12	6	6	Итоговый проект
	Итого	80	32	48	0

2.3. Учебная программа

2.3.1 Введение в анализ данных и машинное обучение

Цель данной программы — предоставить участникам базовые знания и навыки, необходимые для успешного начала работы в области анализа данных и машинного обучения, а также подготовить их к дальнейшему углубленному изучению этих дисциплин.

Понятия анализа данных: основы, цели и методы анализа данных, включая описательную, диагностическую и прогностическую аналитику.

Инструменты: обзор популярных инструментов для анализа данных, таких как Python, Jupyter Notebook.

NumPy: изучение библиотеки для работы с многомерными массивами данных, выполнение математических операций и обработка больших объемов информации.

Pandas: освоение библиотеки для манипуляции данными, базовые функции Pandas.

2.3.2. Работа с данными

Данный модуль охватывает ключевые аспекты обработки и подготовки данных для анализа и машинного обучения. Участники познакомятся с основными концепциями, методами и инструментами, необходимыми для эффективной работы с данными. Модуль содержит следующие темы:

Знакомство с данными: источники и типы данных

Участники узнают о различных источниках данных (CSV, Excel, базы данных, API) и различиях между структурированными и неструктурированными данными. Обсуждаются основные форматы хранения данных.

Очистка данных: техники и инструменты

В этом разделе рассматриваются методы обработки пропусков, дубликатов, выбросов и аномалий. Участники изучат использование библиотеки Pandas для очистки данных, а также визуализацию с помощью Matplotlib и Seaborn. Также будут рассмотрены методы масштабирования данных и создание новых признаков.

Преобразование данных: техника и практика

Участники изучат основные методы преобразования данных, включая нормализацию, стандартизацию и кодирование категориальных переменных. Будут представлены примеры использования функций Pandas и NumPy для этих задач.

Подготовка данных для анализа и машинного обучения

В этом разделе участники научатся создавать наборы для обучения и тестирования, выбирать значимые признаки и обрабатывать текстовые и временные данные.

2.3.3. Основы машинного обучения

Модуль "Основы машинного обучения" предоставляет участникам фундаментальные знания о ключевых алгоритмах машинного обучения и их применении, а также развивает практические навыки работы с данными. Это позволит участникам уверенно применять полученные знания в своих проектах и исследованиях.

Лекционные занятия

Алгоритмы обучения с учителем: линейные модели для регрессии и классификации.

Участники изучат основы линейных моделей, их применение для задач регрессии и классификации, а также методы оценки их эффективности.

Алгоритмы обучения с учителем: деревья решений и их применение

Рассматриваются принципы работы деревьев решений, их преимущества и недостатки, а также примеры применения в различных задачах.

Методы классификации и регрессии: за пределами линейных и деревьев решений

Обучение без учителя: кластеризация и методы понижения размерности

В этом разделе рассматриваются методы кластеризации (например, K-средние) и техники понижения размерности (например, PCA), которые помогают анализировать данные без предварительной разметки.

Практические занятия

Практическая часть включает в себя применение изученных алгоритмов на реальных данных с использованием библиотек Python (таких как Scikit-learn). Участники будут работать над проектами, где смогут реализовать алгоритмы машинного обучения, оценить их эффективность и визуализировать результаты.

2.3.4. Продвинутые методы машинного обучения

Модуль "Продвинутые методы машинного обучения" включает 4 академических часа лекций и 8 академических часов практики, направленных на углубленное изучение методов оптимизации и алгоритмов машинного обучения.

Лекционные занятия

Управляемая оптимизация: Алгоритмы и подходы

Участники изучат основные алгоритмы оптимизации, включая градиентный спуск и его вариации, а также адаптивные методы. Будут рассмотрены условия сходимости и механизмы оптимизации.

Оптимизация гиперпараметров

Рассмотрение методов, таких как сеточный поиск, случайный поиск и байесовская оптимизация для настройки гиперпараметров моделей.

Принципы работы ансамблей

Участники познакомятся с методами объединения нескольких слабых моделей в одну сильную. Обсуждаются алгоритмы бустинга (например, AdaBoost, Gradient Boosting) и случайные леса, включая CatBoost.

Регуляризация моделей
Обсуждение методов регуляризации для улучшения обобщающей способности моделей. Будут приведены примеры применения регуляризации и обсуждены ее преимущества.

Практические занятия
Реализация и настройка случайных лесов
Участники научатся настраивать параметры случайных лесов и оценивать их эффективность на реальных данных.

Работа с градиентным бустингом
Практика применения алгоритмов градиентного бустинга для решения задач классификации и регрессии.

Примеры регуляризации моделей
Участники изучат различные методы регуляризации, применяя их на практике для повышения качества моделей.

Модуль "Продвинутые методы машинного обучения" предоставляет участникам углубленные знания о современных методах оптимизации и алгоритмах машинного обучения. Освоенные техники помогут участникам эффективно применять данные методы в своих проектах, улучшая качество и точность моделей.

2.3.5. Введение в нейронные сети

Модуль "Введение в нейронные сети" включает 6 академических часов лекций и 10 академических часов практики, направленных на ознакомление участников с основами нейронных сетей и их применением в задачах машинного обучения.

Лекционные занятия
Основы обучения нейронных сетей
Участники изучат принципы работы нейронных сетей, алгоритмы обучения и оптимизации, включая градиентный спуск.

Введение в библиотеку Keras и TensorFlow
Обзор библиотек Keras и TensorFlow, их структуры и возможностей для построения нейронных сетей.

Введение в глубокие нейронные сети (DNN)
Рассмотрение архитектуры глубоких нейронных сетей и их применения в различных задачах.

Сверточные нейронные сети (CNN): основы и применение. Рекуррентные нейронные сети (RNN): основы и применение

Трансформеры

Введение в архитектуру трансформеров с использованием библиотеки Hugging Face, демонстрация создания модели с помощью минимального количества кода.

Практические занятия

Создание и настройка нейронной сети с использованием Keras

Участники научатся строить и настраивать простые нейронные сети.

Практические задания по обучению и оптимизации моделей

Участники выполнят задания по обучению различных моделей и оптимизации их параметров.

Модуль "Введение в нейронные сети" предоставляет участникам необходимые знания о базовых концепциях и методах работы с нейронными сетями, а также развивает практические навыки их применения. Освоенные техники помогут участникам понимать устройство нейронных сетей.

2.3.6. Итоговый проект

Модуль "Итоговый проект" включает в себя лекции и практические занятия, направленные на подготовку участников к реализации и защите своих проектов.

Лекции

Обсуждение итогового проекта, представление тем

В ходе лекции участники обсудят цели и задачи итогового проекта, а также представят свои идеи и темы для работы. Обсуждение поможет определить актуальность выбранных тем и их соответствие требованиям курса.

Пример оформления и реализации

Лекция посвящена методам оформления проектов и основным принципам их реализации. Участники получат рекомендации по структуре проекта, оформлению документации и представлению результатов.

Подготовка к защите проекта

В этой лекции будут рассмотрены ключевые аспекты подготовки к защите проекта, включая создание презентации, формулирование основных выводов и ответов на возможные вопросы.

Практические занятия

Поиск и подбор тем

Участники самостоятельно исследуют возможные темы для своих проектов, анализируя актуальные проблемы и подходы в области анализа данных и машинного обучения.

Подготовка проекта

Участники работают над структурированием своих проектов, собирая необходимые данные, выбирая методы анализа и разрабатывая план реализации.

Примеры готовых проектов

Участники ознакомятся с примерами успешных проектов, что поможет им лучше понять требования и стандарты оформления.

2.3.7. Проекты

Помимо многочисленных проверочных и самостоятельных проектов, в Программе предусмотрены специальный итоговый проект, для обобщения полученных знания. На нем особое внимание уделяется освоению практического применения полученных знаний на реальных проектах.

Раздел 3. «Итоговый проект»

Итоговый проект является важной частью курса "Введение в анализ данных и машинное обучение" и предоставляет учащимся возможность применить полученные знания на практике. Проект включает в себя несколько ключевых этапов, которые помогут студентам глубже понять процесс анализа данных и машинного обучения.

Цели итогового проекта

- Применение теоретических знаний: участники должны продемонстрировать умение применять алгоритмы и методы, изученные в курсе, для решения реальных задач.
- Разработка практического решения: на основе анализа данных студенты создадут проект, который может включать в себя разработку модели машинного обучения или анализ набора данных.
- Оценка результатов: итоговый проект позволит оценить уровень готовности учащихся к практическому применению методов анализа данных.

Структура итогового проекта

- Выбор темы: участники выбирают тему проекта, которая соответствует их интересам и учебным целям. Темы могут варьироваться от анализа открытых данных до разработки моделей предсказания.
- Сбор и предобработка данных: участники выбирают проект из предложенных тем.
- Анализ данных: на этом этапе студенты применяют различные методы анализа данных, используя библиотеки Python (например, Pandas, NumPy) для визуализации и интерпретации результатов.
- Моделирование: участники разрабатывают модели машинного обучения с использованием алгоритмов, изученных в курсе (например, линейная регрессия,

деревья решений, SVM). Они также оптимизируют гиперпараметры моделей для достижения наилучших результатов.

- Подготовка отчета и презентации: участники готовят отчет о выполненной работе, который включает описание процесса, результаты анализа и выводы. Презентация должна быть структурированной и содержать визуальные материалы (графики, таблицы).

- Защита проекта: участники представляют свои проекты перед преподавателями и однокурсниками, отвечая на вопросы и получая обратную связь.

Оценивание итогового проекта

- Оценка проектов будет производиться на основе следующих критериев:
- Актуальность и оригинальность темы
- Качество сбора и предобработки данных
- Методы анализа и моделирования
- Результаты и выводы
- Презентация и защита проекта

Заключение

Итоговый проект по курсу "Введение в анализ данных и машинное обучение" позволяет учащимся интегрировать теоретические знания с практическими навыками. Успешное выполнение проекта не только повысит уверенность студентов в своих способностях, но и подготовит их к дальнейшему обучению или профессиональной деятельности в области анализа данных.

Раздел 4. «Организационно-педагогические условия реализации программы»

4.1. Учебно-методическое обеспечение и информационное обеспечение программы (литература)

Нормативно-правовые документы:

1. Федерального закона «Об образовании в Российской Федерации (от 29 декабря 2012 г. № 273-ФЗ).
2. Приказа Минобрнауки России от 1 июля 2013 г. № 499 «Об утверждении Порядка организации и осуществления образовательной деятельности по дополнительным профессиональным программам».
3. Постановления Правительства Российской Федерации от 11 октября 2023 г. № 1678 «Об утверждении Правил применения организациями, осуществляющими образовательную деятельность, электронного обучения, дистанционных образовательных технологий при реализации образовательных программ».
4. Федерального государственного образовательного стандарта высшего образования по направлению подготовки 09.03.02 Информационные системы и технологии, утвержденный приказом Минобрнауки России от 19 сентября 2017 г. N 926 (с изменениями и дополнениями от 26 ноября 2020 г., 8 февраля 2021 г.)
5. Профессионального стандарта «06.042 Специалист по большим данным», утвержденный приказом Министерства труда и социальной защиты Российской Федерации от 06.07.2020 года №405н;

4.2. Методические и учебные материалы

Каждое занятие Программы описано в методическом пособии для преподавателей. Общий алгоритм проведения занятий состоит из нескольких этапов:

- постановка целей урока, описание того, что учащиеся должны достигнуть в результате урока;
- указание на место урока и его функции в общей системе обучения;
- демонстрационное решение типовых (модельных) заданий по теме урока, с выделением алгоритма решений такого рода задач;
- закрепление у учащихся сформулированного алгоритма.

В процессе урока учащимися применяется учебное пособие, разработанное ООО «Юмакс».

4.3. Материально-технические условия реализации программы

Образовательные технологии, используемые в процессе реализации программы:

Операционная система, браузер, антивирусная программа, программа видеосвязи, компилятор кода, программа для работы с электронными таблицами.

4.4. Педагогические условия реализации программы (требования к преподавательскому составу)

К реализации программы привлечены представители компаний со стажем работы в области искусственного интеллекта и в смежных областях, имеющие опыт решения практических задач с использованием технологий искусственного интеллекта от 3 лет в течение последних 10 лет в профильной компании или в профильном подразделении, а также не выполняют функции иностранного агента.

К реализации программы привлечены представители компаний со стажем работы в области искусственного интеллекта и в смежных областях, имеют опыт решения практических задач с использованием технологий искусственного интеллекта от 3 лет в течение последних 10 лет в профильной компании или в профильном подразделении, а также не выполняют функции иностранного агента.

Основная литература:

1. "Введение в анализ данных" Б. Г. Миркин. Издательство: Юрайт. 2015 г.
2. "Машинное обучение без лишних слов". Андрей Бурков. Издательство: Питер. 2020 г.
3. "Прикладное машинное обучение с помощью Scikit-Learn и TensorFlow". Орельен Жерон. Издательство: Питер. 2018 г.
4. "Глубокое обучение". Иошуа Бенджио, Ян Гудфеллоу, Аарон Курвилль. Издательство: Питер. 2019 г. (перевод)
5. "Машинное обучение и TensorFlow". Нишант Шакла. Издательство: Питер. 2020 г.
6. "Идеи машинного обучения". Шай Шалев-Шварц, Шай Бен-Давид. Издательство: Вильямс

Дополнительная литература:

1. Долин В.Б. Python для начинающих. М.: ДМК Пресс, 2018.

2. Соломатин М.В. Программирование на Python 3. СПБ.: БХВ-Петербург, 2019.
3. Кольцов М.С. Python. СПБ.: Питер, 2019.
4. Лутц М. Программирование на Python. Том 1. М.: Символ-Плюс, 2011.
5. Васильев А.Н. Программирование на Python в примерах и задачах. М.: Эксмо, 2021.

Дополнительные интернет-ресурсы:

https://www.python.org/doc/
https://docs.python.org/3/tutorial/inputoutput.html
https://docs.python.org/3/howto/functional.html
https://docs.python.org/3/tutorial/classes.html
https://docs.python-telegram-bot.org/en/latest/
https://git-scm.com/
https://www.djangoproject.com/
https://html.com/html5/