TECHNICAL
EDUCATION
ACADEMY

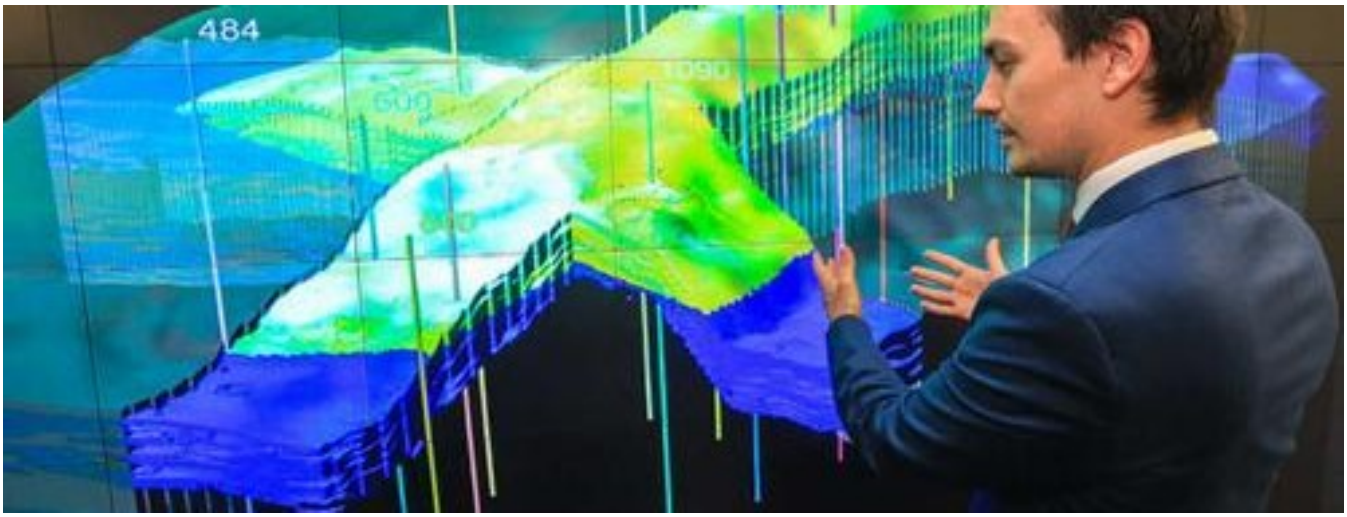
Школа инжиниринга

ПРОГРАММА КУРСА

000268

GEOP019 – Расширенный Анализ Данных и Машинное Обучение для интегрированных подповерхностных решений

Специалисты по нефти и газу ежедневно работают с огромным объемом информации.



Учись у Лучших в Своем Деле!

О программе

Продвинутый уровень

Зачем этот курс

Специалисты по нефти и газу ежедневно работают с огромным объемом информации. Эта информация необходима для построения точных статических и динамических моделей коллектора. Однако значительная часть информации остается неиспользованной человеческой «погрешностью» и из-за ограниченных вычислительных ресурсов. Эта недостающая часть может оказать значительное влияние на точность модели коллектора. Понимание машинного обучения и методов Big Data может помочь наиболее полно использовать входную информацию и достичь наиболее точных результатов.

Этот курс объясняет и дает рекомендации более опытным участникам о том, как эффективно применять методы машинного обучения в наших реальных проектах. Он построен для объяснения теории методов машинного обучения (регрессии, деревьев решений, нейронных сетей, опорных векторных машин, ансамблей алгоритмов), предлагает рекомендации по практическому использованию этих методов и стратегии для наиболее подходящего выбора метода.

Этот 10-дневный курс более практичен, всем специалистам будут представлены все инструменты и алгоритмы для создания моделей машинного обучения. Это отличная возможность получить практический опыт для сбора и подготовки входных данных, применяя методы машинного обучения

Для кого и в чём польза

Геологоразведчики и инженеры-геофизики, инженеры-геодезисты, специалисты по геостатистике и профессионалы, желающие понять концепции машинного обучения, применять методы машинного обучения в практических задачах, совершенствовать свои аналитические и интерпретирующие навыки

Чему вы научитесь

- Анализировать, перепроизводить и визуализировать большие наборы данных в Python
- Понимать основы машинного обучения и будут способны решать проблем с использованием методов машинного обучения
- Использовать алгоритмы машинного обучения для всестороннего анализа данных и смогут строить точные предсказающие модели
- Эффективно сотрудничать со специалистами по обработки данных в команде проекта

Содержание программы

Модули 01–05

01 Введение

- Знакомство с курсом и зачем его проходить
- Как устроена специализация анализа данных и машинного обучения
- Знакомство с Python и его установка
- Знакомство и работа с IPyton Notebook
- Синтаксис Python
- Библиотеки Python (Pandas, NumPy, Scipy, Matplotlib)

02 Повторение базовых понятий математического анализа и основ линейной алгебры для понимания теории методов машинного обучения

- Производная и градиент
- Векторы, Матрицы и операции с ними

03 Повторение базовых концепций теории вероятностей и статистики, которые необходимы для понимания механизма работы всех методов анализа данных

- Вероятность и случайная величина
- Распределение, плотность и свойства распределений
- Основные формулы теории вероятности
- Виды распределений (нормальное, Бернулли)

04 Знакомство с методами оптимизации

- Метод градиентного спуска
- Метод Ньютона

05 Введение в машинное обучение

- Основные термины в машинном обучении
- Типы задач в машинном обучении (обучение с учителем (supervised learning), задачи классификации и регрессии, обучение без учителя (unsupervised learning), clustering, anomaly detection, dimensionality reduction)
- Метрики качества и функции потерь

Содержание программы

Модули 05–08 · Модуль 05 — продолжение

- Регуляризация
- Проблема переобучения: bias-variance tradeoff
- Кросс-валидация (cross-validation)
- Знакомство с библиотекой scikit-learn для анализа петрофизических данных
- Обзор open source data community (Kaggle, etc)

06 Линейная и логистические регрессии

- Подготовка данных для обучения (создание и проверка признаков, очистка данных)
- Визуализация данных
- Lab: подготовка данных для обучения
- Линейная регрессия
- Lab: линейная регрессия на петрофизических данных
- Логистическая регрессия
- Lab: логистическая регрессия на петрофизических данных

07 Деревья решений и их композиции

- Основы деревьев решений
- Lab: построение дерева решений
- Random Forest
- Lab: тренировка Random Forest на петрофизических данных
- Градиентный бустинг
- Lab: тренировка композиции деревьев решений на петрофизических данных с использованием библиотеки XGBoost

08 Нейронные сети

- Основы нейронных сетей
- Backpropagation
- Lab: тренировка нейронной сети на петрофизических данных с использованием библиотеки Keras
- Deep Learning и другие типы нейронных сетей (autoencoder, recurrent neural networks, convolutional neural networks, generative adversarial networks)

Содержание программы

Модули 09–14

09 Метод опорных векторов (SVM)

- Основы метода опорных векторов
- Понятие ядра
- Lab: тренировка SVM на петрофизических данных

10 Обучение без учителя

- Методы кластерного анализа (на примере метода K-Means)
- Метод главных компонент (PCA)
- Смесь нормальных распределений (Mixture of Gaussians)

11 Рекомендации и заключение

- Как подготовить «сырые» данные к обучению и создать новые признаки?
- Как выбрать лучший алгоритм?
- Как настроить алгоритм и выбрать лучший набор параметров?
- Как оценить качество работы алгоритма?

12 Байесовские модели

- Байесовские модели

13 Подробнее о методах кластеризации

- Подробнее о методах кластеризации

14 Обзор статей по машинному обучению при решении геолого-седиментологических задач

- Обзор статей по машинному обучению при решении геолого-седиментологических задач

Предварительные требования

- Чтобы участники могли полностью освоить продвинутое содержание этого курса и максимально использовать свой учебный опыт, рекомендуется наличие следующих предварительных знаний:

Содержание программы — продолжение

Профессиональный опыт:
Опыт в отрасли: Не менее 2–3 лет работы в нефтегазовой промышленности, особенно на должностях геолога, петрофизика, инженера по пластам или геостатистического моделера.
Знания о недрах: Знакомство с типами подземных данных, концепциями моделирования пластов, процессами разведки и разработки.

Навыки программирования:
Владение Python: Базовые или средние навыки программирования на Python, включая опыт работы с библиотеками, такими как NumPy, pandas, matplotlib и seaborn для обработки и визуализации данных.
Скриптинг и автоматизация: Умение писать скрипты для автоматизации простых задач и рабочих процессов обработки данных.

Математические основы:
Статистика и теория вероятностей: Понимание базовых статистических понятий, таких как среднее, медиана, стандартное отклонение, дисперсия, корреляция и основная теория вероятностей.
Линейная алгебра и математический анализ: Знание векторов, матриц, дифференцирования и интегрирования, которые являются фундаментальными для алгоритмов машинного обучения.

Навыки анализа данных:
Обработка данных: Опыт работы с большими наборами данных, очисткой данных и методами предварительной обработки.
Методы визуализации: Способность создавать и интерпретировать базовые визуализации данных для выявления закономерностей и инсайтов.

Базовое понимание машинного обучения (рекомендуется, но не обязательно):
Концептуальное понимание: Знание фундаментальных концепций машинного обучения, таких как обучение с учителем и без учителя, задачи регрессии и классификации.
Знакомство с алгоритмами: Базовые знания о распространенных алгоритмах, таких как линейная регрессия, решающие деревья или методы кластеризации.

Программное обеспечение и инструменты:
Среда разработки: Знакомство с использованием интегрированных сред разработки (IDE) таких, как Jupyter Notebook или PyCharm.

Содержание программы — продолжение

Системы контроля версий (опционально): Базовое понимание контроля версий с использованием Git для целей совместной работы.

Мягкие навыки:
Аналитическое мышление: Сильные аналитические навыки и умение решать проблемы для эффективной интерпретации данных и результатов.
Командная работа: Готовность эффективно сотрудничать с специалистами по данным и междисциплинарными командами.

Примечание: Участникам рекомендуется освежить свои знания в перечисленных областях перед началом курса, чтобы обеспечить гладкий и продуктивный учебный процесс. Тем, у кого отсутствуют некоторые из предварительных знаний, следует рассмотреть возможность предварительного обучения или самообразования для формирования необходимой базы.

Участие в программе

Чтобы уточнить даты, формат проведения, стоимость и условия участия, обратитесь в Академию и укажите код курса 000268.

Обучение для команды

При обращении укажите задачи предприятия, должности участников, размер группы, предпочтительные сроки и место проведения. Эти сведения помогут обсудить подходящий вариант обучения.

Подготовка к участию

Перед записью ознакомьтесь с разделом «Предварительные требования» в этой программе.

Обсудить обучение

mail@tecedu.org
www.tecedu.academy

Учись у Лучших в Своем Деле!