TECHNICAL
EDUCATION
ACADEMY

Школа инжиниринга

ПРОГРАММА КУРСА

000224

GEOP009 - 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

Средний 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.

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, препроизводить 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 предсказывающие модели.
- Эффективно сотрудничать со специалистами по обработки данных в команде проекта.

Содержание программы

Модули 01–03

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 основ математического анализа, линейной алгебры, теории вероятности и статистики, а также методов оптимизации для понимания механизма работы всех методов анализа данных

- Производная и градиент
- Векторы, Матрицы и операции с ними
- Вероятность и случайная величина
- Распределение, плотность и свойства распределений
- Основные формулы теории вероятности
- Доверительные интервалы
- Метод градиентного спуска
- Метод Ньютона

03 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
- 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)

Содержание программы

Модули 04–07

04 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

05 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

06 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)

07 Обучение без учителя

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

Содержание программы

Модули 08–09

08 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?

---

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

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

Участие в программе

Чтобы уточнить даты, формат проведения, стоимость и условия участия, обратитесь в Академию и укажите код курса 000224.

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

В исходной программе предварительные квалификационные требования не указаны. Необходимую подготовку можно уточнить перед записью.

Обсудить обучение

mail@tecedu.org
www.tecedu.academy

Учись у Лучших в Своем Деле!