Utility Analytics 201:

Applied Machine Learning for Utility Professionals

About this Training

Description

Today, widespread Machine Learning (ML) in the Utility industry has become more prevalent than imagined a few short years ago. This training introduces ML, including the concepts of exploratory data analysis, supervised learning (classification and regression), and unsupervised learning (dimensionality reduction and clustering). Participants will apply concepts through guided, hands-on activities.

Audience

This training is intended for the following audiences:

1. Analytics professionals who are interested in machine learning methods with applications in utilities.
2. *Utility Analytics 101* completers who want to continue advancing their in-depth knowledge of analytics in the utilities setting.

Prerequisites

College- or university-level statistics and algebra or equivalent experience. Some exposure to statistical programming (for example, in Python or R language) is helpful but not required.

Objectives

Upon completion of this training, participants will be able to:

1. Provide an overview of machine learning and related tools and topics.
2. Apply exploratory analysis, supervised learning, and unsupervised learning techniques towards industry use cases.
3. Be able to use predictive analytics methods to produce insights or solutions to a problem, given appropriate datasets and tools.
4. Understand how to evaluate and improve models and perform error detection/correction.

Certification

Participants who register and complete the course will earn a Certificate of Participation and Completion and Continuing Education Units (CEUs) from the University of Oklahoma.

Utility Analytics 201:

Applied Machine Learning for Utility Professionals

Course Topical Outline

Day 1

Introduction: Welcome and Overview

- Welcome: introductions
- Welcome: expectations
- Welcome: goals and feedback
- Welcome: software overview

Brief Machine Learning Overview

- Definition
- Relationship to artificial intelligence, machine learning, deep learning, data science
- Applications: why use machine learning?
- Applications: real-world examples
- Overview: problem types (Supervised Learning, Unsupervised Learning, Reinforcement Learning)
- Overview: batch vs. online learning
- Overview: Instance- vs. model-based
- Overview: trends (R&D, technical performance, economy & education)

Preparing for Machine Learning

- ML Process Model
- Business understanding: business understanding and performance metrics
- Exploratory analysis: data types and characteristics, data visualization for exploratory analysis, outliers, missing data
- Data preparation: pre-processing and feature engineering
- Data Modeling: training and performance assessment
- Visualization and communication

Hands-on Application: preparing data for machine learning

Day 1 Wrap-up: Recap and day 2 preview

Day 2

Supervised learning

- Conceptual overview
- Important algorithms: linear regression, logistic regression, decision tree learning, ensembles, support vector machines, k-Nearest Neighbors, MLP neural networks
- Anatomy of a learning algorithm: building blocks (loss function, optimization criterion, optimization routine), gradient descent

Assessing performance

- Regression: RMSE

- Classification: confusion matrix, precision/recall, accuracy, AUC

Improving performance

- Hyperparameter tuning
- Regularization

Learning algorithm selection: guiding questions and algorithm peculiarities

Hands-on Application: supervised learning (classification)

Day 2 Wrap-up: Recap and day 3 preview

Day 3

Advanced methods

- Ensembles: boosting and bagging
- Artificial Neural Networks: MLPs

Challenges in Machine Learning

- Insufficient quantity of data
- Non-representative training data
- Poor-quality data
- Irrelevant features
- Underfitting
- Overfitting

Ethical concerns and identifying biases

Hands-on Application: supervised learning (regression)

Unsupervised methods

- Conceptual overview
- Important algorithms: PCA, k-means clustering, hierarchical (agglomerative) clustering, density-based clustering

Time Series Modeling

- Time series patterns, white noise, stationarity (non-stationarity), time series forecasting

Hands-on Application: unsupervised learning

Day 3 Wrap-up: Recap and next steps