

N479: Applied Statistical Modeling and Big Data Analytics

Instructor(s): Srikanta Mishra

Format and Duration

Classroom - 3 Days

Virtual - 5 Sessions

Summary

Business Impact: This training course will provide a hands-on introduction to **statistical modeling** and **big data analytics** so that participants can use them for **petroleum engineering** and **geoscience applications**.

Topics to be covered include: (a) easy-to-understand descriptions of the commonly-used techniques, (b) case studies demonstrating the applicability, limitations and value-added proposition for these methods, and (c) hands-on problems sessions using open source and/or commercial software. This course will provide engineers and geologists with practical techniques for identifying hidden patterns and relationships in large datasets and extracting data-driven insights towards actionable information that can contribute to lower cost, improved efficiency and/or increased productivity in oil and gas operations. This class will arm petroleum engineers and geoscientists with advanced capabilities to extract new insights from E&P data that can help: (a) learn hidden patterns and relationships in geologic datasets, (b) identify production sweet spots in developed plays; (c) determine factors responsible for separating good wells from poor producers wells, (d) build fast surrogate models of reservoir performance, and (e) assist in predictive maintenance by identifying failure inducing conditions from historical records.

For a more in depth summary of D479 please use the following link to watch Dr. Mishra discuss his course in detail:

<https://vimeo.com/427050656/aa5159e905>

Learning Outcomes

Participants will learn to:

1. Apply foundational concepts in probability and statistics for basic data analysis
2. Perform linear regression for building simple input-output models
3. Conduct multivariate data reduction and clustering for finding sub-groups of data that have similar attributes
4. Converse with confidence about big data, data analytics and machine learning terminology and fundamental concepts, and critically review topical technical publications
5. Apply machine learning techniques for regression and classification for developing data-driven input-output models
6. Evaluate proxy modeling and uncertainty quantification studies for probabilistic performance forecasting

Training Method

This is a classroom or virtual classroom consisting of lectures interspersed with worked examples, hands-on exercises and discussions.

The textbook for the course will be “Applied Statistical Modeling and Data Analytics: A Practical Guide for the Petroleum Geosciences” by Srikanta Mishra and Akhil Datta-Gupta (Elsevier, 2017), supplemented by

N479: Applied Statistical Modeling and Big Data Analytics

Instructor(s): Srikanta Mishra

Format and Duration

Classroom - 3 Days

Virtual - 5 Sessions

various technical publications.

Who Should Attend

This course is designed for petroleum engineers, geoscientists, and managers interested in becoming smart users of statistical modeling and data analytics.

Prerequisites and Linking Courses

Participants should have a basic knowledge of statistics or should have attended N480 (Introduction to Statistical Modeling & Big Data Analytics).

Course Content

(1) Foundational Concepts

- Big data technologies, basic data analytics and machine learning terminology/concepts
- Data, statistics, and probability
- Distributions (models, fitting distributions to data)
- Inference (Confidence limits, bootstrap, significance tests)
- Data visualization
- Problem session

(2) Basic Regression Analysis

- Linear regression (univariate and multivariate regression)
- Understanding regression statistics, ANOVA
- Non-parametric regression
- Problem session

(3) Multivariate Data Analysis

- Dimension reduction (Principal component analysis)
- Cluster analysis (K-means, hierarchical clustering, self-organizing maps)
- Problem session

(4) Machine Learning Basics

- Overview of techniques
- Evaluating model performance (model validation, goodness-of-fit, common pitfalls)
- Variable importance
- Model aggregation
- Case studies

(5) Machine Learning for Regression and Classification I

N479: Applied Statistical Modeling and Big Data Analytics

Instructor(s): Srikanta Mishra

Format and Duration

Classroom - 3 Days

Virtual - 5 Sessions

- Classification/regression trees
- Random forest
- Gradient boosting machine
- Problem session

(6) Machine Learning for Regression and Classification II

- Support vector machine
- Neural networks and deep learning
- Problem session

(7) Miscellaneous topics and Wrap-up

- Experimental design and response surface analysis
- Uncertainty quantification
- Selected literature review
- Key takeaways and resources
- Data analytics do's and don't's