

---

## N480: Introduction to Statistical Modeling and Big Data Analytics

Format and Duration  
Classroom - 1 Day

Instructor(s): Srikanta Mishra

---

### Summary

This training course will provide an introduction to statistical modeling and big data analytics for petroleum engineering and geoscience applications. Topics to be covered include: (a) easy-to-understand descriptions of the commonly-used techniques, and (b) case studies demonstrating the applicability, limitations and value-added proposition for these methods. This course will inform engineers and geologists about techniques for data-driven analysis that can convert data into actionable information for reducing cost, improving efficiency and/or increasing productivity in oil and gas operations.

### Learning Outcomes

Participants will learn to:

1. Apply foundational concepts in probability and statistics for basic data analysis
2. Interpret linear regression for building simple input-output models
3. Examine 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
5. Differentiate machine learning techniques for regression and classification for developing data-driven input-output models
6. Critique uncertainty quantification studies for probabilistic performance forecasting

### Training Method

A one-day classroom course consisting of lectures with worked examples.

### Who Should Attend

This course is for designed for petroleum engineers, geoscientists, and managers interested in learning about the basics of statistical modeling and data analytics.

### Course Content

1. Big data technologies, basic data analytics and machine learning terminology/concepts
2. Exploratory data analysis, probability distributions, confidence limits
3. Basic linear regression
4. Data reduction, cluster analysis and data visualization
5. Machine learning basics, techniques for regression and classification problems
6. Machine learning case studies
7. Uncertainty quantification
8. Wrap-up