Department of Chemical Engineering
Indian Institute of Technology Guwahati
Guwahati, Assam-781039, India

A 3 Day Workshop On

Applied Statistical Modeling and Data Analytics for Petroleum Engineering and Related Applications

Instructor: Dr. Srikanta Mishra, Honorary Faculty, Chemical Engineering, IIT Guwahati

Course Content *(Topics 1-5 include lectures, worked problems, hands-on exercises, discussion)*

(1) Foundational Concepts (Day 1)

- 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)

(2) Basic Regression Analysis (Day 1)

- Linear regression (univariate and multivariate regression)
- Understanding regression statistics
- Non-parametric regression

(3) Multivariate Data Analysis (Day 1)

- Dimension reduction (Principal component analysis)
- Cluster analysis (K-means, Hierarchical clustering)
- Data visualization

(4) Machine Learning Basics (Day 2)

- Overview of techniques
- Evaluating model performance (model validation, goodness-of-fit, common pitfalls)
- Ensemble modeling (aggregation)
- Variable importance

(5) Machine Learning for Regression and Classification (Day 2)

- Tree-based methods (decision trees, Random forest, Gradient boosting machine)
- Non-tree-based methods (Neural network, Support vector machine)

(6) Group Projects (Day 3)

Participants will be divided into multi-disciplinary groups, and will develop and present a machine learning model as a capstone project

(7) Wrap-up (Day 3)

- Key takeaways and resources
- Data analytics do's and don'ts