IBM approved MachineLearning

Machine Learning with Python

	What will you learn at the end of this training? • Python programming basics & Data Handling in Python • Basic Statistics and Reporting • Machine Learning tools and techniques • Basic and advanced algorithms in Machine Learning • Model building and Validation techniques
	Pre-requisites for the training • Minimum BSc/ BTech • Basic mathematical skills (12th class mathematics) • Knowledge on statistics is **NOT** necessary **Training Details and Features** • 100% hands on sessions • Duration –24 Hours – 3Days • Learning through problem solving and case studies

	Trainer Profile • Venkata Reddy Konasani • 10+ years – Data Scientist / 5+ years – Corporate Training • Conducted 5000+ hours training and 60+ corporate batches • Author of the book "Practical Business Analytics using SAS" • Rich industry experience as applied Data Analyst and Data Scientist • Experience in credit risk model building, market response model building, social media analytics, revenue forecasting and machine learning • Post Graduate in Applied Statistics and Informatics from IIT Bombay • https://statinfer.com/venkat-profile/

Course Curriculum

Day-1: Python Programming Basic Statistics & Regression

1. 1. Introduction to Machine Learning
    - a. Introduction to Machine Learning
    - b. Machine Learning tools and Techniques
    - 2. Data Handling, Data Validation and Graphs
    - a. Important packages used in Machine Learning
    - b. Data importing
    - c.

    - Working with datasets
    - d. Descriptive statistics
    - e. Central Tendency
    - f. Variance
    - g. Percentiles
    - h. Outlier detection
    - i. Variable distribution charts
    - 3. Regression Analysis
    - a. Correlation
    - b. Simple Regression models

- c. R-Square
- d. Multiple regression
- e. Multicollinearity
- f. Individual Variable Impact
- g. Air passenger's data case study
- h. SAT score data case study

Day-2: Classification using Logistic Regression and Trees

- • Logistic Regression
- o Need of logistic Regression
- o Logistic regression models
- o Validation of logistic regression models
- o

- Multicollinearity in logistic regression
- o Individual Impact of variables
- o Confusion Matrix
- o Service Provider Attrition data case study
- • Decision Trees
- o Segmentation
- o Entropy
- o Information gain
- o Building Decision Trees
- o Validation of Trees
- o Pruning the trees
- o Fine tuning the trees
- o Prediction using Trees
- o Customer retention case study
- • Cluster Analysis
- o Supervised vs unsupervised learning
- o Need of Cluster Analysis
- o K- Means clustering algorithm
- o The theory behind cluster Analysis
- o Building and interpreting clusters

Day-3: Model Validation Techniques and Neural Networks

- • Model Selection and Cross validation
- o How to validate a model?
- o What is a best model?
- o Types of data
- o Types of errors
- o The problem of over fitting
- o The problem of under fitting
- o Bias Variance Trade-off
- o Cross validation techniques

- Neural Networks
- Neural network Intuition
- Neural network and vocabulary
- Neural network algorithm
- Math behind neural network algorithm
- Building the neural networks
- Validating the neural network model
- Neural network applications
- Image recognition using neural networks
- Course Conclusion and Q&A
- Couse conclusion
- Reference books, videos and blogs
- Next steps
- Final Q&A
- Fina assessment (optional)

Appendix

List of Case Studies used in the course

1. Online Sales data case study – Data Handling in Python
2. USA Census Income data analysis – Basic statistics
3. Air passenger prediction and driver analysis -Regression
4. SAT score prediction and driver analysis -Regression
5. E-com Website sales prediction case study -Regression
6. Product sales analysis – Logistic Regression
7. Customer attrition analysis -Logistic Regression
8. Customer Survey Segmentation and Drivers – Decision Trees
9. Internet service provider customer segmentation – Decision Trees
10. Retail customer segments – Cluster Analysis
11. Customer attrition analysis – Model selection and cross validation
12. Productivity data -Neural networks
13. Image recognition -Neural networks