MACHINE LEARNING

By Dr. Vishwanath Rao

Data Science and Machine Learning Training with Python– 5Days

- ? Machine Learning Introduction
- ? Statistics vs Business Analytics vs Data Science vs Machine Learning vs Deep
  Learning vs Artificial Intelligence(Understanding the difference)
- ? Machine learning project life cycle
- ? Roles and skill set for Data Science/Machine Learning
- ? Which role I can play in Machine learning Project?
- ? Machine learning architecture
- ? Generalized architecture
- ? Tools and platforms used in Machine learning

o Cloud based platforms
o Proprietary tools
o Open source tools, Platforms
? Introduction Analytics Tool(Python) (Day-1) o What is Python & History
o Installing Python & Python Environment o Basic commands in Python
o Data Types and Operations
o Python packages
o Loops
o My first python program o If-then-else statement
o Functions in Python
o User defined Functions
? Important libraries for Data Science o Numpy

o Scipy
o Pandas
o Matplotlib o Sklearn

[?] Data Handling in Python o Data importing
(Day-1)
o Connecting to External data sources o Working with datasets
o Manipulating the datasets
o Exporting the datasets into external files o Data Merging
[?] Basic Descriptive Statistics (Day-2) o Population and Sample
o Data Types
o Measures of Central tendency o Measures of dispersion
o Percentiles & Quartiles
o Box plots and outlier detection o Creating Graphs and Reporting o Probability Distributions
o Hypothesis testing
[?] Data Preparation for Analysis (Day-2)
o Exploratory Data Analysis o Data Validation rules
o Data Cleaning techniques
o Data Preparation for analysis

- [?] Deal with missing data
- [?] Add default values
- [?] Remove incomplete rows
- [?] Deal with error-prone columns
- [?] Fixing the nan values and string/float confusion
- [?] Normalize data types
- [?] Change casing
- [?] Creating new variables
- [?] Feature Scaling
- [?] Feature Standardization
- [?] Label Encoding
- [?] One-Hot Encoding

🈳 Algorithms used in Machine Learning(Day-3) o Supervised Machine learning algorithms
o Unsupervised Machine learning algorithms
🈳 Regression Analysis (Day-3) o Correlation
o Simple Regression models o R-Square
o Multiple regression
o Multicollinearity
o Individual Variable Impact o Case Study
o Shrinkage methods

- 🈳 Logistic Regression (Day-3)
  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 Case study

- 🈳 Decision Trees (Day-3)

o Types - Classification and Regression trees o Gini Index
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 case study
🈳 Model Selection and Cross validation (Day-4) 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 Tradeoff
o Cross validation
o Boot strapping
o House price index data case study o case study
🈳 Neural Networks(Day-4)
o Neural network Intuition
o Neural network and vocabulary
o Neural network algorithm
o Math behind neural network algorithm
o Building the neural networks –

o Perceptron .Multi layer NN, Backpropagation o Validating the neural network model
o Neural network applications
o Image recognition using neural networks
⍰ Random Forest and Boosting(Day-5) o Introduction
o Ensemble Learning
o How ensemble learning works
o Bagging
o Building models using Bagging
o Random Forest algorithm
o Random Forest model building
o Finetuning hyper parameters
o Variable Importance plots
o Marketing Reponses model building for e-mail marketing
⍰ Text Mining and NLP(Day-5) o What is text mining
o The NLTK package
o Preparing text for analysis o Text summarisation
o Sentiment analysis
o Naïve bayes technique
o Text classification
o News data classification
o Topic Modelling
o LDA
o LDA on Python
Small end to end Case Study – Use case
⍰ Machine Learning Project (Day-5) o The issue of large data
o The issue of many variables o ML best practices