GAMAKA
Artificial Intelligence Solutions

Office No 309, Paranjape – The Business Hub, Karve Road, Kothrud, Pune - 411038
Email: itplacements@gamakaai.com Cell: 91-7378483656. WhatsApp: 91-7378493293
www.gamakaai.com

List of Projects/Case Studies

1. Forecasting Stock and Commodity Prices
2. Analysing Wine Types and Quality
3. Build your own image recognition model with TensorFlow
4. Customer Segmentation and Effective Cross Selling
5. Predict fraud with data visualization & predictive modelling
6. Analysing Movie Reviews Sentiment
7. Analysing Music Trends and Recommendations
8. Spam Detection
9. Build your own Recommendation System
10. Build your own Python predictive modelling, regression analysis & machine learning Model
11. Football Players (Estimating Population Mean from a Sample)
12. Election Polling (Estimating Population Proportion from a Sample)
13. A Medical Study (Hypothesis Test for the Population Mean)
    Employee Behavior (Hypothesis Test for the Population Proportion)
14. A/B Testing (Comparing the means of two populations
15. Customer Analysis (Comparing the proportions of 2 populations)
16. Predictive medicine: prognosis and diagnostic accuracy
17. Virtual assistance for patients and customer support
18. Creation of drugs - allows choosing, which experiments should be done and incorporates all the new information in a continuous learning loop
19. Clustering algorithms for customer segmentation
20. Discovering similarities across my Spotify music using data, clustering and visualization
21. An End-to-End Project on Time Series Analysis and Forecasting with Python
22. Using LSTMs to forecast time-series
23. Evolution of a salesman: A complete genetic algorithm tutorial for Python
24. A Machine Learning Approach — Building a Hotel Recommendation Engine
25. How To Create Data Products That Are Magical Using Sequence-to-Sequence Models

GAMAKA
Artificial Intelligence Solutions

Office No 309, Paranjape – The Business Hub, Karve Road, Kothrud, Pune - 411038
Email: itplacements@gamakaai.com Cell: 91-7378483656. WhatsApp: 91-7378493293
www.gamakaai.com

Module 2: Data Science – R-Programming

Duration: 90 Hours with hands on tutorials

12 Case Studies with Internship

Mode of training: Online/Offline(classroom)/Combo (Offline & Online)

Introduction to Data Science

- What is Data Analytics?
- Types of Data Sets and Data Models
- Understanding of Business Analytics
- Need of Business Analytics
- Types of Business Analytics
- Descriptive Analytics
- Predictive Analytics
- Prescriptive Analytics
- Supply Chain Analytics
- Health Care Analytics
- Marketing Analytics
- Human Resource Analytics
- Data Management and Business Analytics
- Web Analytics and Business Intelligence
- Data Science as a Strategic Asset
- Data Warehousing and OLAP
- Data Visualization using R and Excel
- Data Visualization using Tableau BigData and Data Science

Fundamentals of R

- Understanding R
- Which Companies Use R?
- Understanding Comprehensive R Archive Network (CRAN)
- How to Install R on Operating Systems?
- How to Install R on Windows from CRAN Website?
- IDEs for R
- R Packages: Installation and Practice
- Understanding R Programming
- Studying Operators in R
- Operators: Arithmetic, Relational, Logical, Assignments
- Statements in R Programming
- Conditional Statements in R
- Break and Next Statement
- If else () Function
- Switch Function
- Scan () Function
- Loops in R
- How to Run an R Script and Batch Script?
- R Functions: Commonly Used and String Functions

R Data Structures

- Defining Data Structures in R
- Types of Data Structures
- Vectors and Scalars
- Colon Operator
- Matrices
- Elements: Vector, Matrix, Array
- Understanding Data Frames
- Factors and Lists
- How to Import Files in R?
- How to Import Minitab File?
- Importing Table and CSV Files
- Importing Data from SQL Databases How to Export Files from R?
- Types of Apply Functions
- Apply () Function: Lapply, Sapply, Tapply
- Vapply () Function, Mapply () Function

derstanding Dplyr Package

Office No 309, Paranjape – The Business Hub, Karve Road, Kothrud, Pune - 411038
Email: itplacements@gamakaai.com Cell: 91-7378483656. WhatsApp: 91-7378493293
www.gamakaai.com

- How to Import an Excel File?

Supervised Learning Techniques and the Implementation of Algorithms

- R Data Structures – Vectors, Factors, Lists, Data Frames, Matrixes and Arrays
- Managing Data with R
- Saving and Loading R Data Structures
- Importing and Saving Data from CSV Files
- Importing Data from SQL Databases
- Exploring the Structure of Data
- Exploring and Understanding Data
- Exploring Numeric Variables
- Understanding Types of Data
- Qualitative and Quantitative Analysis Studying Descriptive Statistics
- Exploring Numeric Variables
- Measuring the Central Tendency – The Model Measuring Spread – Variance and Standard Deviation
- Visualizing Numeric Variables – Boxplots and Histograms
- Understanding Numeric Data – Uniform and Normal Distributions
- Measuring the Central Tendency – The Mode Exploring Relationships between Variables
- Visualizing Relationships – Scatterplots Nominal and Ordinal Measurement
- Interval and Ratio Measurement
- Statistical Investigation
- Inferential Statistics
- Probability and Central Limit Theorem Exploratory Data Analysis
- Normal Distribution
- Distance Measures
- Euclidean & Manhattan Distance
- Cosine
- Correlation
- PPMC (Pearson Product Moment Coorelation)

Hypothesis Testing

- Importance of Hypothesis Testing in Business
- Null and Alternate Hypothesis
- Understanding Types of Errors
- Contingency Table and Decision Making
- Confidence Coefficient
- Upper Tail Test and Test Statistics
- Understanding Parametric Tests
- Z-Test and Z-Test in R
- Chi-Square Test
- Degree of Freedom
- One-Way ANOVA Test
- F-Distribution, F-Ration Test

Introduction to Data Science and Machine Learning

- Implementing Machine Learning Algorithms on larger Data Sets with Apache Mahout
- How do Machines Learn?
- Abstraction and Knowledge Representation
- Generalization
- Assessing the Success of Learning
- Steps to apply Machine Learning to your Data
- Choosing a Machine Learning Algorithm
- Thinking about the Input Data
- Thinking about Types of Machine Learning Algorithms
- Matching your Data to an Appropriate Algorithm

Machine Learning Techniques using R

- Data Preparation for Modelling
- How do Machines Learn?
- Choosing a Machine Learning Algorithm
- Clustering
- Similarity Metrics
- Distance Measure Types: Euclidean

Office No 309, Paranjape – The Business Hub, Karve Road, Kothrud, Pune - 411038
Email: itplacements@gamakaai.com Cell: 91-7378483656. WhatsApp: 91-7378493293
www.gamakaai.com

- Machine Learning Techniques Using R
- Machine Learning: Tasks, Features, Models, and Design
- Machine Learning Common Use Case
- Supervised and Unsupervised Learning Techniques
- Cosine Measures
- Creating predictive Models
- Classification using Nearest Neighbors

Supervised Learning Techniques and the Implementation of Various Algorithms

- Supervised Learning Techniques and Algorithms
- Understanding Process Flow of Supervised Learning Techniques
- K-NN, Naïve Bayes, Support Vector Machines
- Defining Classification
- Understanding Classification and Prediction
- Decision Tree Classifier
- How to Build Decision Trees?
- Basic Algorithm for a Decision Tree
- Decision Trees and Data Mining
- Random Forest Classifier
- Features of Random Forests
- Out of Box Error Estimate and Variable Importance
- Naïve Bayes Classifier Model
- Bayesian Theorem
- Advantages and Disadvantages of Naïve Bayes Classifier Model
- Understanding Support Vector Machines
- What is Geometric Margin SVMs?
- Understanding Linear SVMs

Unsupervised Machine Learning Techniques – Implementation of Different Algorithms

- Studying Clustering
- Clustering and Classification
- Understanding K-means Clustering
- K-means and Pseudo Code
- K-means Clustering using R
- TF-IDF and Cosine Similarity
- Application to Vector Space Model
- What is Hierarchical Clustering?
- Hierarchical Clustering Algorithm
- Understanding Agglomerative Clustering Process
- DBSCAN Clustering
- What is Association Rule Mining?
- Association Rule Strength Measures
- Checking Apriori Algorithms
- Ordering Items
- Understanding Candidate Generation
- Performing Visualization on Associated Rules

Regression Methods for Forecasting Numeric Data

- What is Regression?
- Model Selection
- Generalized Regression
- Simple Linear Regression
- Multiple Linear Regression
- Correlations
- Prediction: Time Dependent/Variant Data
- Ordinary Least Square Regression Model
- Dummy Variable Regression Model
- Interaction Regression Model
- Non-Linear Regression Model

Office No 309, Paranjape – The Business Hub, Karve Road, Kothrud, Pune - 411038
Email: itplacements@gamakaai.com Cell: 91-7378483656. WhatsApp: 91-7378493293
www.gamakaai.com

- Correlation between X and Y
- Ridge and Regularized Regression
- LASSO
- Time Series
- Perform Regression Analysis with Multiple Variables
  Non-Linear and Linear Models

Deep Learning – Neural Networks and Support Vector Machines

- Understanding Neural Networks
- From Biological to Artificial Neurons
- Activation Functions
- Network Topology
- Neural Networks: Master Feed-Forward
- Recurrent and Gaussian Neural Network
- The Number of Layers
- The Direction of Information Travel
  The Number of Nodes in Each Layer
- Training Neural Networks with Backpropagation
- Support Vector Machines
- Classification with Hyperplanes
- Finding the Maximum Margin
- The Case of Linearly Separable Data
- The Case of Non-Linearly Separable Data
- Retrieve Data using SQL Statements
- Using Kernels for Non-Linear Spaces
- Performing OCR with SVM