
*www.bbds.ma*

Mo.medwani@bbds.ma

# BBDS 13 Week Data Science Training Program

**Empower Data Science Teams with the knowledge, skills, and confidence to successfully create & implement reproducible data products & predictive models**.

BBDS's Data Science Course will prepare you for the job role of a data scientist and will help you gain data scientist skillset by learning data science using analytical tools like Python, R, RapidMiner, SAS, Tableau, BigML, Watson Analytics.

The program is an equivalent of a Master Degree in Data Science + 4 years of experience. It will help you master analytical techniques like data exploration, data visualization and various predictive analytic techniques by implementing real-life, industry-oriented data science projects using Python/R data science programming languages.

The program will also help you gain expertise about various popular machine learning algorithms like Decision Trees, K-mean Clustering, Gradient Boosting, Boosted Trees, Random Forest, and Naïve Bayes using Python/R programming languages.

The program is best suited for beginners and also experienced professionals who would like to use Python/R for doing data science.

# About BBDS !


mo.medwani@bigbang-datascience.com
678-209-9780

# BBDS Team

## Nathan Bumham

- ✓ Nathan Burnham is a Computer Engineer with extensive deep learning experience who also founded a robotics startup
- ✓ 4 Years experience in Machine Learning and Robotics
- ✓ 7 Years experience with Python
- ✓ Nathan taught Deep Learning as a Masters class in 2017 & 2016

## Suresh Gilakamsettii

- ✓ Currently working as a Big Data Engineer with MNC
- ✓ Industry Certified Big Data Developer
- ✓ 5 years of IT industry experience
- ✓ Worked for Google, EY, Nissan, TD Bank etc.
- ✓ Trained more than 1000 associates on Big Data and Analytics
- ✓ Published a paper on “Time Series Classification using kernel weighted k-NN and DTW algorithms” in an International conference

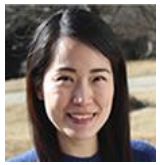
## Julie Grantier

- ✓ Taught statistics and computer programming, including SAS, SPSS, and
- ✓ Consulting on data science and research design projects in a variety of fields
- ✓ Providing training in statistics and computer programming including R, Visual Basic, and Java

## Shan Nabi

- ✓ 14+ Years of experience in IT
- ✓ 7+ Years of experience in SAP consultant in Cloud Services
- ✓ 1+ Years of experience in Data Science and related technologies
- ✓ Master degrees (MS-IT)

## Jingjing Cannon



- ✓ Data Scientist in a Telecommunication Company
- ✓ 6+ years in data science and statistical modeling
- ✓ Last year PhD student at Georgia State University majoring in Computational Neuroscience
- ✓ Specialties: Advanced Data Science, Text Mining, NLP, Behavioral Analysis

## Mo Medwani

- ✓ 15+ Years of experience in IT (Service Delivery Management)
- ✓ 7+ Years of experience in Data Analytics
- ✓ 3+ Years of experience in Data Science and related technologies
- ✓ 3 Master degrees (MBA, MS-IT & MS- Data Science)
- ✓ Ph.D. Candidate in Data Science
- ✓ Founder of Big Bang Data Science Solutions

## Muhammad Junaid

- ✓ Over 25 years Experience in Computer, Business, Sales & Marketing
- ✓ Two Bachelors [Business & Computers: Specialized In E-Commerce]
- ✓ Diploma Achievements from Computer Electronics & Computer Science

## Edward Bujak

- ✓ 25 years of experience in IT
- ✓ 18 years of experience in education: computer science, mathematics, engineering
- ✓ 2 Masters degrees (MS-Electrical Engineering & Computer Science, MS-Education

# BBDS Programs?


Data Science End to End
(13 Weeks Program)


Data Science with Python
(8 Weeks Program)

Deep Learning
(8 Weeks Program)


Advanced R-Bootcamp
(6 Weeks Program)


Big Data & Hadoop
(8 Weeks Program)


NLP
(6 Weeks Program)

Data Visualization

List of Big Bang Data Science Institute training services provided

http://www.bbds.ma

# Assumption & Certificates


mo.medwani@bigbang-datascience.com
678-209-9780

# Assumption & Industry Corticate in line with !

- **The course Assumption**

The course assume that you know **close to nothing about Data Science and ML**. Its goal is to give you the concepts, intuitions you need to actually implement programs capable of learning from data

We will cover large number of techniques, from simplest and most commonly used (such as Linear Regression) to some Deep Learning techniques that regularly win competitions

- **Industry Certificates**

At the end of the program, you should be ready for CAP & EMCDSA Certificates

EMC Data Scientist Associate (EMCDSA) Certification.

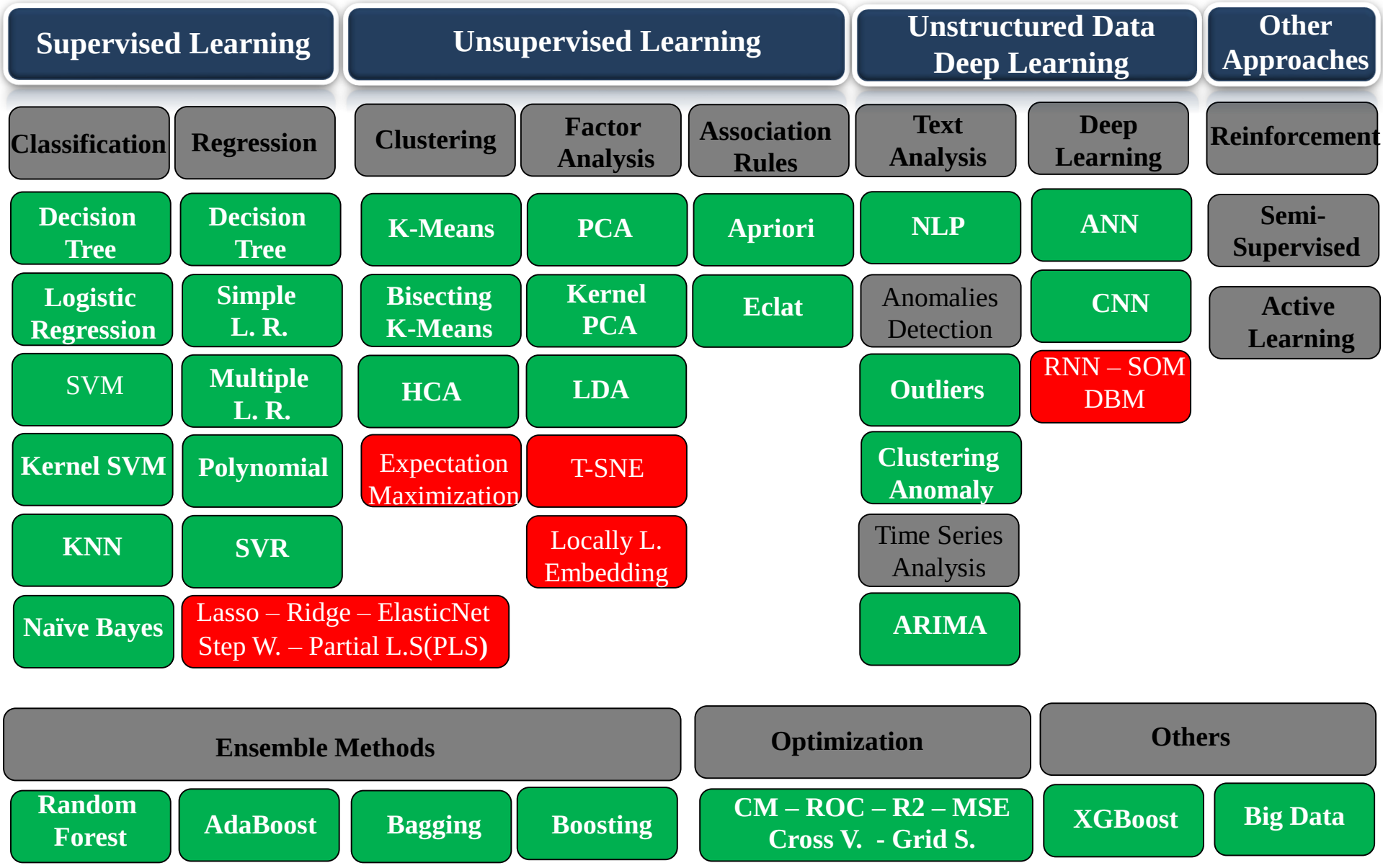
CAP: Certified Analytics Professional from INFORMS

# Machine Learning Techniques Covered !


mo.medwani@bigbang-datascience.com
678-209-9780

# Machine Learning - Artificial Intelligence

Machine Learning techniques covered on the 13 Week Program



# Weekly Program Structure


mo.medwani@bigbang-datascience.com
678-209-9780

# Sessions Schedule

| SESSIONS | MEETINGS | DAYS | TIME | INSTRUCTOR |
|---|---|---|---|---|
| Session 1 | Online/Onsite | Saturdays | 3:00 PM to 6:00 + PM EST | Mo Medwani |
| Session 2 | Online | Tuesdays | 8:00 PM to 11:00 + PM EST | Mo Medwani |
| Session 3 | Online | Thursdays | 8:00 PM to 11:00 + PM EST | Mo Medwani |
| Session4 | Online | Fridays | 8:00 PM to 11:00 + PM EST | Tanvir R. Shaik + Mo Medwani |

Mondays, Wednesdays: Available for extra 1:1 sessions and recaps if needed

Sundays – OFF

# Course Schedule & Structure


## Week 1 - Data Science Fundamentals

Learn Data Science basics, Good understanding of Data Science and Data Analytics, Overview of EMC and CAP certificates, Introduction to concepts, methodologies and best practices

**Session 1:** Introduction to Data Science - CRISP-DM

- Data Science Process - CRISP-DM
- Projects Discussion (Milestone Projects Assignment)

**Session 2:** Crash Course & hands-on in R

- R & RStudio - Fundamentals Working Directory - Assigning Objects
- Data Structures (Vectors, Factors, Lists, Data frames, Matrix, Matrices)
- Functions – Packages - Comments

**Session 3:** Crash Course & hands-on in Python

- Jupyter Notebook & Notebook Azure - Arithmetic & Assignment
- Lists - Dictionaries – Sets – Tuples – Functions – Lambda - Conditions
- Five Most common Python Packages

**Session 4:** Data Visualization and Analytical tools

- Data Viz with Tableau, Seaborn, Ggplot, Matplotlib
- Data Science with RapidMiner, BigML, SAS, IBM Watson
- Case studies with real-world business problems

# Course Schedule & Structure


## Week 2 - Business & Problem Understanding

Determine Business Objectives, Assess Situation, and Determine Data Mining Goals, Produce Project Plan. Good understanding of problem framing, Decision Framing, Decision Analysis and Decision implementation using Decision First Molder

**Session 1:** Decisions Management & Problem Framing

- Decisions First - Quantifying Business Problem - Identifying Y variable
- Measure of Success – Stakeholders - Project Plan

**Session 2:** Crush Course in Statistics

- Descriptive Statistics - Probability Theory - Statistical Inference
- Chi-Square Test - Correlation – T.Test - ANOVA - MANOVA

**Session 3:** Data Loading & Manipulation in (R & Python)

- Importing Libraries - Installing packages - Reading Data - Dataset subletting
- Data Manipulation - Data Munging

**Session 4:** Data Visualization and Analytical tools

- Data Viz with Tableau, Seaborn, Ggplot, Matplotlib
- Data Science with RapidMiner, BigML, SAS, IBM Watson
- Case studies with real-world business problems

# Course Schedule & Structure

## Week 3 - Data Understanding & Data Preparation

Exploratory Data Analysis using R & Python, Descriptive statistics, hypothesis testing, data preprocessing, missing values imputation, data transformation, Dive deep into R programming language from basic syntax to advanced packages and data visualization (e.g. reshape2, dplyr, string manipulation, ggplot2, R Shiny).

**Session 1:** Data Preparation & Processing

- Types of Data - Data Quality

**Session 2:** EDA & Data Visualization - (R & Python)

- Statistical Analysis (John T. 5 Numbers Summary)
- Correlation - Variance & Covariance - Skewness & Kurtosis.
- Missing Values & Outliers - Duplicates & redundant
- Data Visualization (Histograms, Bar Plot, Scatter Plot, Box Plot)
- R & Python tutorials with real world business problems

**Session 3:** Feature Selection & Dimensionality Reduction - (R & Python)

- Aggregation - Sampling - Dimensionality Reduction
- Feature Subset Selection - Feature Creation - Discretization & Binarization
- Variable Transformation - Imbalancing Data - Data Partitioning

**Session 4:** Data Visualization and Analytical tools

- Data Viz with Tableau, Seaborn, Ggplot, Matplotlib
- Data Science with RapidMiner, BigML, SAS, IBM Watson
- Case studies with real-world business problems

# Course Schedule & Structure

## Week 4 - Predictive Analytics – Classification (Non-Parametric)

Deepen machine learning skills with R and scikit learn. Focus on data cleaning, feature extraction modeling, and model selection using Supervised Learning

**Session 1:** Decision Tree Classification in (R & Python)

- A Basic Algorithm to Build a Decision Tree
- Methods for Expressing Attribute Test Conditions
- Measures for Selecting an Attribute Test Condition
- Algorithm for Decision Tree Induction
- Characteristics of Decision Tree Classifiers - Strengths & Weaknesses
- R & Python tutorials with real world business problems

**Session 2:** Rule-Based Classifier in (R & Python)

- How a Rule-Based Classifier Works - Properties of a Rule Set
- Direct Methods for Rule Extraction - Indirect Methods for Rule Extraction
- Characteristics of Rule-Based Classifiers - Strengths & Weaknesses
- R & Python tutorials with real world business problems

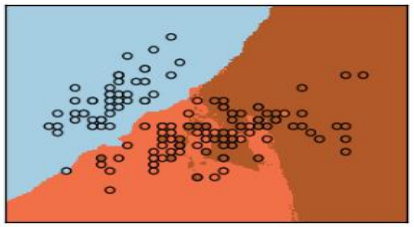
**Session 3:** Naïve Bayes - (R & Python)

- Basics of Probability Theory - Naïve Bayes Assumption - Strengths & Weaknesse
- R & Python tutorials with real world business problems

**Session 4:** Data Visualization and Analytical tools

- Data Viz with Tableau, Seaborn, Ggplot, Matplotlib
- Data Science with RapidMiner, BigML, SAS, IBM Watson

# Course Schedule & Structure



## Week 5 - Predictive Analytics – Classification (Parametric)

Deepen machine learning skills with R and scikit learn. Focus on data cleaning, feature extraction, modeling, and model selection using Supervised Learning

**Session 1:** Logistic Regression in (R & Python)

- Logistic Regression as a Generalized Linear Model
- Learning Model Parameters - Characteristics of Logistic Regression
- Strengths & Weaknesses
- R & Python tutorials with real world business problems

**Session 2:** Support Vector Machine (SVM) & KSVM in (R & Python)

- Margin of a Separating Hyperplane - Linear SVM - Soft-margin SVM
- Nonlinear SVM - Characteristics of SVM - Strengths & Weaknesses
- R & Python tutorials with real world business problems

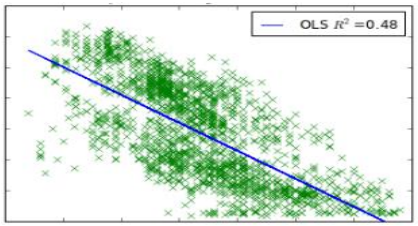
**Session 3:** Neighbor Classifiers (KNN) in (R & Python) non-parametric

- Algorithm - Characteristics of Nearest Neighbor Classifiers
- Strengths & Weaknesses
- R & Python tutorials with real world business problems

**Session 4:** Data Visualization and Analytical tools

- Data Viz with Tableau, Seaborn, Ggplot, Matplotlib
- Data Science with RapidMiner, BigML, SAS, IBM Watson
- Case studies with real-world business problems

# Course Schedule & Structure




## Week 6 - Predictive Analytics – Regression (Non-Linear)

Deepen machine learning skills with R and scikit learn. Focus on data cleaning, feature extraction, modeling, and model selection using Supervised Learning

**Session 1:** Decision Tree Regression - (R & Python)

- A Basic Algorithm to Build a Decision Tree
- Methods for Expressing Attribute Test Conditions
- Measures for Selecting an Attribute Test Condition
- Algorithm for Decision Tree Induction
- Characteristics of Decision Tree Classifiers - Strengths & Weaknesses
- R & Python tutorials with real world business problems

**Session 2:** Support Vector Machine (SVR) - (R & Python)

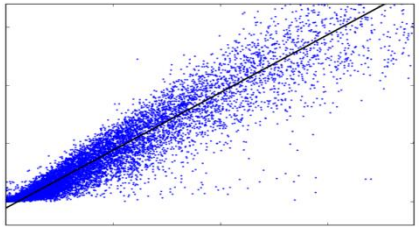
- Margin of a Separating Hyperplane - Linear SVM
- Soft-margin SVM - Nonlinear SVM - Characteristics of SVM
- Strengths & Weaknesses
- R & Python tutorials with real world business problems

**Session 3:** Projects Discussion (Milestone Projects Assignment)

**Session 4:** Data Visualization and Analytical tools

- Data Viz with Tableau, Seaborn, Ggplot, Matplotlib
- Data Science with RapidMiner, BigML, SAS, IBM Watson
- Case studies with real-world business problems

# Course Schedule & Structure



## Week 7 - Predictive Analytics – Regression (Linear)

Deepen machine learning skills with R and scikit learn. Focus on data cleaning, feature extraction, modeling, and model selection using Supervised Learning

**Session 1:** Simple Linear Regression - (R & Python)

- Ordinary Least Squares - Gradient Descent
- Regularization - Making Predictions with Linear Regression
- Preparing Data for Linear Regression
- R & Python tutorials with real world business problems

**Session 2:** Multiple Linear Regression - (R & Python)

- Ordinary Least Squares - Gradient Descent
- Regularization - Making Predictions with Linear Regression
- Preparing Data for Linear Regression

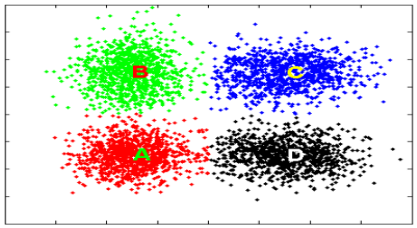
**Session 3:** Polynomial Linear Regression - (R & Python)

- Algorithm - Why use polynomial regression?
- Polynomial Feature ()
- R & Python tutorials with real world business problems

**Session 4:** Data Visualization and Analytical tools

- Data Viz with Tableau, Seaborn, Ggplot, Matplotlib
- Data Science with RapidMiner, BigML, SAS, IBM Watson
- Case studies with real-world business problems

# Course Schedule & Structure




## Week 8 - Descriptive Analysis – Clustering & Anomalies

Deepen machine learning skills with R and scikit learn. Focus on data cleaning, feature extraction, modeling, and model selection using Unsupervised Learning

**Session 1:** K-Mean Clustering - (R & Python)

- The Basic K-means Algorithm - Bisecting K-means
- K-means & Different Types of Clusters
- K-means as an Optimization Problem - Strengths & Weaknesses
- R & Python tutorials with real world business problems

**Session 2:** Agglomerative Hierarchical Clustering - (R & Python)

- Basic Agglomerative Hierarchical Clustering Algorithm
- Specific Techniques - The Lance-Williams Formula for Cluster Proximity
- Key Issues in Hierarchical Clustering - Outliers - Strengths & Weaknesses
- R & Python tutorials with real world business problems

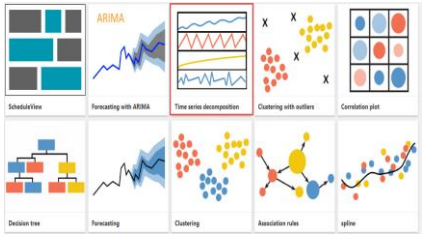
**Session 3:** Anomalies Detection - (R & Python)

- Characteristics of Anomaly Detection Problems & Methods
- Proximity-based Approaches - Strengths & Weaknesses
- R & Python tutorials with real world business problems

**Session 4:** Data Visualization and Analytical tools

- Data Viz with Tableau, Seaborn, Ggplot, Matplotlib
- Data Science with RapidMiner, BigML, SAS, IBM Watson
- Case studies with real-world business problems

# Course Schedule & Structure




## Week 9 - Descriptive Analysis – Association Rules & Time Series

Deepen machine learning skills with R and scikit learn. Focus on data cleaning, feature extraction, modeling, and model selection using NLP, Text Mining, Sentiment Analysis, Deep learning with Theano, TenserFlow & Keras, Neural Networks learn, Convolutional Neural Networks

**Session 1:** Association Rules - (R & Python)

- The Apriori & Eclat Principle
- Frequent Itemset Generation in the Apriori Algorithm
- Candidate Generation & Pruning - Support Counting
- Computational Complexity
- R & Python tutorials with real world business problems

**Session 2:** Times Series Analysis Rules - (R & Python)

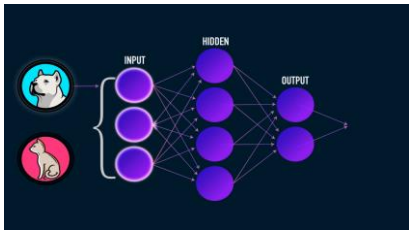
- Forecasting - Relation between time series: Causality & time lags
- Distinction between short & long run - Study of agent's expectations
- Trend removal - Seasonal adjustment
- Detection of structural breaks - Control of the process
- R & Python tutorials with real world business problems

**Session 3:** Project Discussion

**Session 4:** Data Visualization and Analytical tools

- Data Viz with Tableau, Seaborn, Ggplot, Matplotlib
- Data Science with RapidMiner, BigML, SAS, IBM Watson
- Case studies with real-world business problems

# Course Schedule & Structure




## Week 10 - Deep Learning, NLP, Text Mining

Deepen machine learning skills with R and scikit learn. Focus on data cleaning, feature extraction, modeling, and model selection using NLP, Text Mining, Sentiment Analysis, Deep learning with Theano, TenserFlow & Keras, Neural Networks learn, Convolutional Neural Networks

**Session 1:** NLP & Text Analysis - (R & Python)

- Basic feature extraction using text data - Basic Text Pre-processing of text data
- Advance Text Processing
- R & Python tutorials with real world business problems

**Session 2:** Artificial Neural Network (ANN)

- Crash course in TensorFlow, Keras & Theano - Perceptron - Multi-layer Neural Network - Characteristics of ANN
- R & Python tutorials with real world business problems

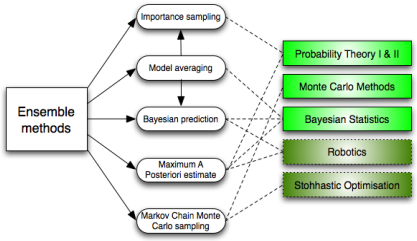
**Session 3:** Deep Learning

- Using Synergistic Loss Functions - Using Responsive Activation Functions
- Regularization - Initialization of Model Parameters
- Characteristics of Deep Learning
- R & Python tutorials with real world business problems

**Session 4:** Data Visualization and Analytical tools

- Data Viz with Tableau, Seaborn, ggplot, Matplotlib
- Data Science with RapidMiner, BigML, SAS, IBM Watson
- Case studies with real-world business problems

# Course Schedule & Structure




## Week 11- Ensemble Methods

Introduction to Cost Function, Object Function, Model Optimization, Model Tuning, Gradient Boosting, Grid and Random Search. Analyze the performance of each algorithms and discuss the result

**Session 1:** Bagging & Boosting & Ada-boost - (R & Python)

- Methods for Constructing an Ensemble Classifier
- Bias-Variance Decomposition - Bagging
- Boosting
- R & Python tutorials with real world business problems

**Session 2:** Random Forest - Voting & Averaging - (R & Python)

- Random Forests - Empirical Comparison among Ensemble Methods
- R & Python tutorials with real world business problems

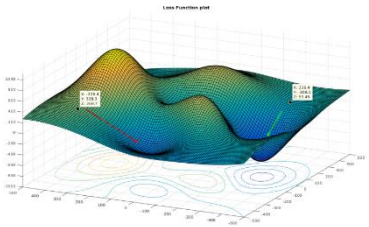
**Session 3:** Class Imbalance Problem

- Building Classifiers - Class Imbalance
- Evaluating Performance - Class Imbalance
- Finding an Optimal Score Threshold
- Aggregate Evaluation of Performance
- R & Python tutorials with real world business problems

**Session 4:** Data Visualization and Analytical tools

- Data Viz with Tableau, Seaborn, ggplot, Matplotlib
- Data Science with RapidMiner, BigML, SAS, IBM Watson
- Case studies with real-world business problems

# Course Schedule & Structure



## Week 12- Model Assessment, Validation, Optimization & Tuning

Introduction to Cost Function, Object Function, Model Optimization, Model Tuning, Gradient Boosting, Grid and Random Search. Analyze the performance of each algorithms and discuss the result

**Session 1:** Model assessments

- CM, ROC, Rank-Ordered Approach - R2, MSE, MAE, Median Error, Median Absolute error, Correlation
- Reasons for Model Overfitting - Model Selection
- Using a Validation Set - Incorporating Model Complexity
- Estimating Statistical Bounds - Model Selection for Decision Trees
- R & Python tutorials with real world business problems

**Session 2:** Model Evaluation

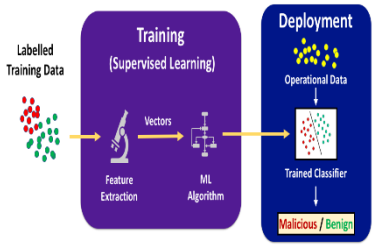
- Holdout Method - Cross-Validation - Presence of Hyper-parameters
- Hyper-parameter Selection - Nested Cross-Validation
- Pitfalls of Model Selection & Evaluation - Overlap between Training & Test Sets
- Use of Validation Error as Generalization Error - Cluster Evaluation
- R & Python tutorials with real world business problems

**Session 3:** Model Comparison

- Estimating the Confidence Interval for Accuracy
- Comparing the Performance of Two Models - XG-boost - (R & Python)
- R & Python tutorials with real world business problems

**Session 4:** Data Visualization and Analytical tools

# Course Schedule & Structure




## Week 13- Model Deployment & False Discoveries

Introduction to Cost Function, Object Function, Model Optimization, Model Tuning, Gradient Boosting, Grid and Random Search. Analyze the performance of each algorithms and discuss the result

**Session 1:** General Deployment Considerations

- Deployment Steps

**Session 2:** Avoiding False Discoveries

- Preliminaries: Statistical Testing
- Modeling Null & Alternative Distributions
- Statistical Testing for Classification
- Statistical Testing for Association Analysis
- Statistical Testing for Cluster Analysis
- Statistical Testing for Anomaly Detection

**Session 3:** Projects submission, Resume & Interview Preparation

# Course Schedule & Structure

## Week 1-13: Capstone Analytical Projects

Complete a capstone project. - Project Selection - Project Scope Analytics Approach - Project Analysis - Data Analysis Techniques Data Analysis Execution Plan- Data Analysis Review - Analytical Technique - Data Model Analysis - Data Analysis Presentation - Final Project Report

- *Lead generation*
- *Program Evaluation*
- *Subscription Churn*
- *Anomaly Detection*
- *Pricing Test*
- *Facebook Ads*
- *Market Basket Analysis*
- *Freemium A/B Testing*
- *Fraud Prevention*

# Why BBDS ?


mo.medwani@bigbang-datascience.com
678-209-9780

# Program Structure

**13 Weeks Program (8th batch stars on Feb 16 to May 20)**

- Frequency 4 Times a week ( 12 to 14 hours a week) – (***Mandatory)***
- 2 to 3 Times a week if needed (***Optional***)
- Extensive Live Online Training
- Instructor-Led Course
- Training Video Recordings
- Quality Training Materials
- Two-Way Interactive Sessions
- Flexible Online Schedules
- Job Oriented Training
- Mock Exam/Assessment
- Graded Assignments & Professional Certificate
- Interview Prep
- Job Placement and Placement Guidance

# The course is designed to enable students to:

- Become an immediate contributor on a data science team
- Assist reframing a business challenge as an analytics challenge
- Deploy a structured lifecycle approach to data analytics problems
- Apply appropriate analytic techniques and tools to analyze big data
- Tell a compelling story with the data to drive business action
- Use open source tools such as R, Python
- Use Tableau, RapidMiner, BigML, SAS for Enterprise Miner, IBM Watson Analytics, IBM BleuMix

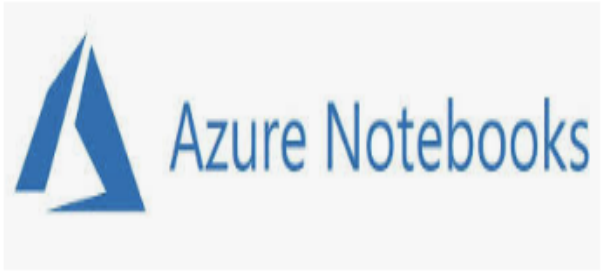
# Languages & Tools





DECISION
MANAGEMENT
SOLUTIONS





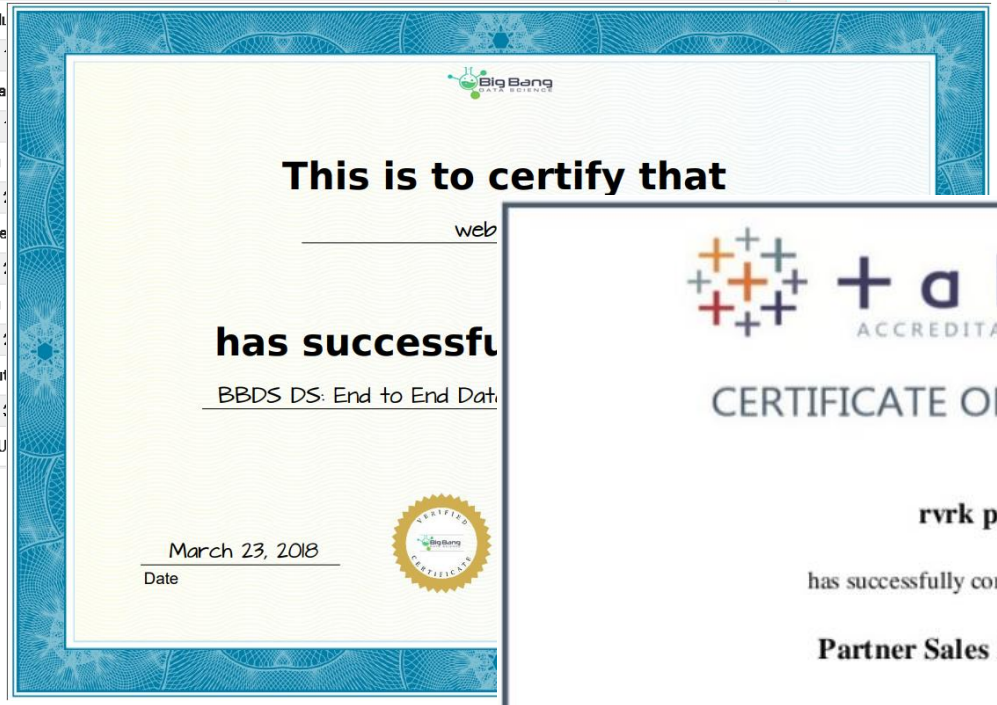

# College Graded Program

- Graded Program

| Course | Progress | Overall Grade | Certificate |
|---|---|---|---|
| BBDS DS: End to End Data Science 13 Week Program | 100% | 98.09% | Download Certificate |

| Module 1 | Course Registration | Quiz Grade |
|---|---|---|
| Unit 1 | Why learning Data Science is an absolute MUST ! | - |
| Module 2 | Week 1 - Session 1 | Quiz Grade |
| Unit 1 | Introdu | |
| Module 3 | Week | |
| Unit 1 | Tablea | |
| Module 4 | Week | |
| Unit 1 | Crash | |
| Module 5 | Week | |
| Unit 1 | Busine | |
| Module 6 | Week | |
| Unit 1 | Crash | |
| Module 7 | Week | |
| Unit 1 | R Input | |
| Module 8 | Week | |
| Unit 1 | Data U | |

- Certificate of Completion





- Tableau Desktop Professional Certificate

# Satisfaction is a Priority

- Complete the program ON TIME (**Show certificate of completion** ) and yet not fully satisfied

# BIG BANG DATA SCIENCE SOLUTIONS

***LEARN . ACHIEVE. STANDOUT***


mo.medwani@bigbang-datascience.com
678-209-9780