CURRICULUM:
DATA SCIENCE PRODEGREE

Greens Technologys

INTRODUCTION - 24 HOURS

BATCH LAUNCH	Intro to Program	Curriculum Overview	Learning Methodology	Guest Lecture
ALL ABOUT DATA	Data	Variables	Data Types	Measures of Central Tendency in Data	Understanding Skewness in Data	Measures of Dispersion	Data Distribution
ANOVA/ REGRESSION ANALYSIS	Analysis of Variance and Covariance	One way analysis of variance	Assumption of ANOVA	Statistics associated with one way analysis of variance	Interpreting the ANOVA Results	Two way analysis of variance	Interpreting the ANOVA Results	Analysis of Covariance	Examine Regression Results	What is Regression Analysis	Linear and Logistic Regression	Statistics Associated with Regression
PREDICTIVE MODELLING	Decision Trees and Neural Networks	Introduction to Predictive Modelling with Decision Trees	Assumptions	Formulate the Model Estimate the Parameters	Check the Prediction Accuracy
TREE AND BAYESIAN NETWORK MODELS	Decision Trees, Bagging	Random Forests, Boosted Trees	Bayesian Classification Models
NEURAL NETWORKS	Perceptron, MLP, Back Propagation	Revision of Key Concepts

R - 66 HOURS

RBASICS	R Base Software	Understanding CRAN	RStudio The IDE	Basic Building Blocks in R	Sequence of Numbers in R	Understanding Vectors in R	Basic Operations Operators and Types
RFUNCTIONS	Handling Missing Values in R	Subsetting Vectors in R	Matrices and Data Frames in R	Logical Statements in R	Lapply, sapply, vapply and tapply Functions
LINEAR REGRESSION THEORY - R	Covariance and Correlation	Multivariate Analysis	Assumptions of Linearity Hypothesis Testing	Limitations of Regression
BUSINESS CASE: MANAGING CREDIT RISK	Business Case : Managing Credit Risk	Meaning of Credit Risk	Impact of Credit Default	Sources of Data for Managing Risk	Understanding Loss Given Default	Understanding Default
LOSS GIVEN DEFAULT LINEAR REGRESSION R	Loss Given Default Linear Regression R	Extract Data in R	Univariate Analysis of Data	Apply Data Transformations	Bivariate Analysis of Data	Identify Multicollinearity in Data	Treatment on Data	Identify Heteroscedasticity Discuss what could be the Reason for Heteroscedasticity	Modelling of Data Variable Significance Identification	Model Significance Test	Predict using Testing Data Set	Validate the Model Performance
LOGISTIC REGRESSION THEORY - R	Reason for Logistic Regression	The Logistic Transform	Logistic Regression Modelling	Model Optimisation	Understanding ROC Curve
PROJECT 1	Project 1 - Default Modelling using Logistic Regression in R
SUPPORT VECTOR MACHINES (THEORY)	Introduction to SVM	Classification as a Hyper Plane Location Problem	Motivation for Linear Support Vectors	SVM as Quadriatic Optimization Problem	Non Linear SVM	Introduction to Kernel Functions
PROJECT 2	Project 2 - Default Modelling using SVM in R
DECISION TREES	Introduction to Decision Trees	Theory of Entropy & Information Gain	Stopping Rules	Overfitting Problem	Cross Validations for Overfitting Problem	Prunning as a Solution for Overfitting	Ensemble Learning Notion	Concept of Bootstrap Aggregation	Concept of Random Forest
BUSINESS CASE	Business Case : Intrusion Detection in IT Network	Meaning of Intrusion in IT Cost of Intrusion	Meaning of Intrusion Detection System
PROJECT 3	Project 3 - Network Intrusion Detection using Decision Tree & Ensemble Learning in R
GUEST LECTURE	Industry View from Expert	Refresher on R	Open House

PYTHON - 35 HOURS

PYTHON BASICS	What is Python?	Installing Anaconda	Understanding the Spyder Integrated Development Environment (IDE)	Lists, tuples, dictionaries, variables
DATA STRUCTURES IN PYTHON USED FOR DATA ANALYSIS	Intro to Numpy Arrays	Creating ndarrays	Indexing	Data Processing using Arrays	File Input and Output	Getting Started with Pandas
DATA FRAME MANIPULATION	Data Acquisition(Import & Export)	Indexing	Selection and Filtering Sorting & Summarizing	Descriptive Statistics	Combining and Merging Data Frames	Removing Duplicates	Discretization and Binning	String Manipulation	PLUS: Project Work on Python
OTHER PREDICTIVE MODELLING TOOLS	Intro to Machine Learning	Random Forests	Sklearn Library & Statsmodels
PROJECT 4	Project 4 - Default Modelling using Logistic Regression in Python
PROJECT 5	Project 5 - Credit Risk Analytics using SVM in Python
PROJECT 6	Project 6 - Intrusion Detection using Decision Trees & Ensemble Learning in Python

SAS - 40 HOURS

INTRODUCTION TO SAS AND SAS PROGRAMS	What is SAS?	Key Features	Submitting a SAS Program	SAS Program Syntax Examining SAS Datasets Accessing SAS Libraries	Sorting and Grouping Reporting Data	Using SAS Formats
READING AND MANIPULATING DATA	Reading SAS Datasets	Reading Excel Data	Reading Raw Files	Reading Database Data	Creating Summary Reports	Combining Datasets
DATA TRANSFORMATIONS	Writing Observations	Writing to Multiple Datasets	Accumulating Total Creating Accumulating Total for a Group of Data	Data Transformations
MACROS	Introduction to Macro Variables	Automatic Macro Variables	User Defined Macro Variables	Macro Variable Reference	Defining and Calling Macros	Macro Parameters	Global and Local Symbol Table	Creating Macro Variables in the Data Step
SQL	Introduction to SQL	How Does RDBMS Work?	SQL Procedures	Specifying Columns	Specifying Rows	Presenting Data	Summarizing Data	Writing Join Queries using SQL	Working with Subqueries, Indexes and Views	Set Operators	Creating Tables and Views using Proc SQL
PROJECT 7	Project 7 - Store Data Analytics in SAS

TABLEAU - 10 HOURS

TABLEAU BASIC	Introduction to Visualization	Working with Tableau	Visualization in Depth Data Organisation	Advanced Visualization	Mapping	Enterprise Dashboards Data Presentation
BEST PRACTICES FOR DASHBOARDING AND REPORTING AND CASE STUDY	Have a Methodology	Know Your Audience	Define Resulting Actions Classify Your Dashboard	Profile Your Data	Use Visual Features Properly	Design Iteratively
INTRODUCTION TO THE GROUP PROJECT	Choice of three projects on various domains

JOB READINESS - 8 HOURS

RESUME BUILDING AND INTERVIEW PREP	Resume Building	Personal Branding	Tips and Resources	Interview Skills
1:1 MOCK INTERVIEWS	1:1 Mock Interviews with Industry Veterans to Clear the Technical Round of Interviews to Give You Confidence to Face Real World Scenarios
GROUP PROJECT PRESENTATION	Groups Present their Project Presentation in Front of Their Peers and industry Experts Evaluate the Solution (Refresher session for online batches)

HANDS-ON PROJECTS

- DEFAULT MODELLING USING LOGISTIC REGRESSION IN R
- DEFAULT MODELLING USING SVM IN R
- NETWORK INTRUSION DETECTION USING DECISION TREE & ENSEMBLE LEARNING IN R
- DEFAULT MODELLING USING LOGISTIC REGRESSION IN PYTHON
- CREDIT RISK ANALYTICS USING SVM IN PYTHON
- INTRUSION DETECTION USING DECISION TREES & ENSEMBLE LEARNING IN PYTHON
- STORE DATA ANALYTICS IN SAS