

-----CURRICULUM-----

BIGDATA & MACHINE LEARNING PRODEGREE

LEVEL 1		INTRODUCTION TO STATISTICS	16 HOURS
BASIC PROBABILITY AND TERMS	Events and their Probabilities Rules of Probability Conditional Probability and Independence Permutations and Combinations Bayers Theorem Descriptive Statistics Compound Probability Conditional Probability	 3 HOURS	
PROBABILITY DISTRIBUTIONS	Types of Distributions Functions of Random Variables Probability Distribution Graphs Confidence Intervals	 3 HOURS	
DATA TRANSFORMATIONS AND QUALITY ANALYSIS	Merge, Rollup, Transpose and Append Missing Analysis and Treatment Outlier Analysis and Treatment	 4 HOURS	
EXPLORATORY DATA ANALYSIS	Summarizing and Visualizing the Important Characteristics of Data Hypothesis Testing Visualizations Univariates, Bivariates Crosstabs, Correlation	 6 HOURS	
LEVEL 2		BASIC OF PYTHON	16 HOURS
INTRODUCTION TO PYTHON	Python Basics Spyder IDE Jupyter Notebook Floats and Strings Simple Input & Output Variables Single and Multiline Comments	 4 HOURS	
CONTROL STRUCTURES	Booleans and Comparisons Conditional Statements (IF ELSE) Operator Precedence Lists - Operations and Functions	 4 HOURS	
FUNCTIONS AND MODULES	Function Arguments Comments and Doc Strings Functions as Objects Modules Standard Lib and Pip	 4 HOURS	


CURRICULUM

BIG DATA & MACHINE LEARNING PRODEGREE


EXCEPTIONS AND FILES	Exception Handling Raising Exceptions Assertions Working With Files	 4 HOURS
LEVEL 3	DATA ANALYSIS WITH PYTHON	32 HOURS
LINEAR REGRESSION	Implementing Simple & Multiple Linear Regression with Python Making Sense of Result Parameters Model Validation Handling Other Issues/Assumptions in Linear Regression: Handling Outliers, Categorical Variables, Autocorrelation, Multicollinearity, Heteroskedasticity Prediction and Confidence Intervals Use Cases	 5 HOURS
LOGISTIC REGRESSION	Implementing Logistic Regression with Python Making Sense of Result Parameters: Wald Test, Likelihood Ratio Test Statistic, Chi-Square Test Goodness of Fit Measures Model Validation: Cross Validation, ROC Curve, Confusion Matrix Use Cases	 5 HOURS
DECISION TREES	Implementing Decision Trees using Python Homogeneity Entropy Information Gain Gini Index Standard Deviation Reduction Vizualizing & Prunning a Tree Implementing Random Forests using Python Random Forest Algorithm Important hyper-parameters of Random Forest for tuning the model Variable Importance Out of Bag Errors	 4 HOURS
PANDAS	Introduction to Pandas IO Tools Basics of NumPy NumPy Functions Pandas - Series and Dataframes	 4 HOURS
SCIKIT LEARN	Introduction to SciKit Learn Load Data into Scikit Learn Run Machine Learning Algorithms Both for Unsupervised and Supervised Data Supervised Methods: Classification & Regression Unsupervised Methods: Clustering, Gaussian Mixture Models Decide What's the Best Model for Every Scenario	 5 HOURS
PROJECT 1	<p>Title: Real Estate Price Prediction using Linear Regression</p> <p>Industry: Real Estate</p> <p>Description: The goal of this Use-case is to make property price predictions using Real Estate data. The dataset contains the of the price of apartments and various characteristics of the property. Based on this data, decide on the price of new properties.</p>	 3 HOURS
PROJECT 2	<p>Title: Bankruptcy Prediction using Logistic Regression</p> <p>Industry: Finance</p> <p>Description: This dataset has various financial ratios based on which a predictive model can be built to predict if a company is going to be bankrupt.</p>	 3 HOURS


CURRICULUM


BIG DATA & MACHINE LEARNING PRODEGREE

PROJECT 3 DECISION TREES	Title : Facebook Post Count using Decision Trees Industry : Social Media Description: Instances in this dataset contain features extracted from facebook posts. The task associated with the data is to predict how many comments the post will receive.	 3 HOURS
--	---	---

LEVEL 4 INTRODUCTION TO MACHINE LEARNING 17 HOURS


INTRODUCTION TO MACHINE LEARNING	What is Machine Learning? End-to-end Process of Investigating Data Through a Machine Learning Lens Evolution and Trends Application of Machine Learning Best Practices of Machine Learning	 3 HOURS
----------------------------------	--	---


MACHINE LEARNING ALGORITHMS	Classification Regression Collaborative Filtering Clustering Principal Component Analysis	 10 HOURS
-----------------------------	--	--


NEURAL NETWORKS	Understanding Neural Networks The Biological Inspiration Perceptron Learning & Binary Classification Backpropagation Learning Learning Feature Vectors for Words Object Recognition	 4 HOURS
-----------------	---	--

LEVEL 5

KERAS	Keras for Classification and Regression in Typical Data Science Problems Setting up KERAS Different Layers in KERAS Creating a Neural Network Training Models and Monitoring Artificial Neural Networks	 4 HOURS
-------	---	---

PROJECT 4 ANN ON KERAS	Title: Credit Default using ANN on Keras Industry: Finance Description: This research aimed at the case of customers' default payments in Taiwan. From the perspective of risk management, the result of predictive accuracy of the estimated probability of default will be more valuable than the binary result of classification - credible or not credible clients.	 4 HOURS
--------------------------------------	--	---

TENSORFLOW	Introducing Tensorflow Neural Networks using Tensorflow Debugging and Monitoring Convolutional Neural Networks Unsupervised Learning	 4 HOURS
------------	--	---

PROJECT 5 CNN ON TENSORFLOW	Title: Handwriting/Facial recognition using CNN on TensorFlow Industry: Pattern Recognition Description: This project will help build a model using Convolutional Neural Network to recognize handwriting/faces	 4 HOURS
---	--	---

CURRICULUM

BIG DATA & MACHINE LEARNING PRODEGREE

LEVEL 6		HADOOP AND SPARK	16 HOURS
INTRODUCTION TO BIG DATA/HADOOP	Big Data and its Sources Popular Tools Used to Store, Process, Analyze & Visualize Big Data RDBMS vs Hadoop Hadoop Architecture and Ecosystem HDFS Design and Architecture Overview When to Use & Not Use Hadoop?	 4 HOURS	
MAPREDUCE FRAMEWORK	MapReduce Design and Execution High level MapReduce Pipeline Strategies for Debugging Hadoop MapReduce Architecture MapReduce I/O Formats	 4 HOURS	
HIVE AND SPARK ECOSYSTEM	Introduction to Hive and Spark Spark Architecture Hadoop vs. Spark Data Sharing in MapReduce Data Types and Validation in Hive Using Hive Built-in Functions Spark Configuration and Cluster Modes	 4 HOURS	
SPARK MACHINE LEARNING LIFECYCLE	Spark Transformers Spark ML Lib Spark ML Pipeline Spark Algorithms for Machine Learning Model Accuracy in Spark	 4 HOURS	
LEVEL 7		THE FUTURE WITH IBM WATSON	15 HOURS
MACHING LEARNING IN 2020	What does the Future of Machine Learning Hold? Virtual Agents Deep Learning Platforms Biometrics Robotics Process Automation	 5 HOURS	
IBM WATSON DEVELOPER	Fundamentals of IBM Watson Advantages of IBM Watson Use cases of Cognitive Services Applications on IBM Watson Administering Watson Applications	 5 HOURS	
PROJECT 6 IBM WATSON	<p>Application on Watson</p> <p>Title: Watson Deep Learning</p> <p>Description: Initiate and monitor batch training experiments then compare cross-model performance in real-time. Efficiently automate searching your network's hyperparameter space to ensure the best model performance with the fewest training runs.</p>	 5 HOURS	
RESUME WRITING	The why, the what and the how of Resumes Personal Branding Tips and Resources Interview Skills Using Social Media CV Discussion	 4 HOURS	



Greens Technologys