upGrad | iiitb

PG Diploma in Data Science

(Program Curriculum)

For Prep Sessions + Batch Start Dates:
Please refer to upgrad.com

Note: This curriculum is subject to change based on inputs from IIITB and Industry

COURSE	MODULE NAME	DESCRIPTION	SESSION	SEGMENT
PREPARATORY COURSE	**Introduction to Python**	*Sharpen your Data Analysis skills with Python, which is the choice of language for simplicity, readability and quick deployment.*	*Understanding UpGrad Coding Console*	Introduction Understanding Primary Actions Understanding Statuses & Important Pointers
			Data Structures in Python	Introduction Getting Started - Installation Introduction to Jupyter Notebook The Basics Lists Tuples Dictionaries Sets
			Control Structures and Functions	Introduction If-Elif-Else Loops Comprehensions Functions Map, Filter, and Reduce
	Python for Data Science	*Learn to clean and manipulate the data using Python's powerful data analysis libraries - NumPy and Pandas.*	*Introduction to NumPy*	Introduction NumPy Basics Creating NumPy Arrays Structure and Content of Arrays Subset, Slice, Index and Iterate through Arrays Multidimensional Arrays Computation Times in NumPy and Standard Python Lists
			Operations on NumPy Arrays	Introduction Basic Operations Operations on Arrays Basic Linear Algebra Operations
			Introduction to Pandas	Introduction Pandas Basics Indexing and Selecting Data Merge and Append Grouping and Summarizing Dataframes Lambda function & Pivot tables
			Getting and Cleaning Data	Introduction Reading Delimited and Relational Databases Reading Data From Websites Getting Data From APIs Reading Data From PDF Files Cleaning Datasets
	Data Analysis using SQL	*Learn the basic and advanced concepts of SQL and add another language to your programming toolkit!*	*Basics of SQL*	An introduction to RDBMS and SQL Basics of SQL Data Retrieval with SQL Compound Functions and Relational Operators Pattern Matching with Wildcards Basics of Sorting Session Summary
			Advanced SQL	Order by Clause Aggregate Functions Group by Clause Having Clause Nested Queries Inner Join Multi Join Outer Join Summary
	Advanced SQL	*Learn the advanced concepts of SQL and gain mastery over this programming language.*	*Database design*	Introduction Defining Data Warehouse Structure of Data Warehouse OLAP vs. OLTP Star Schema How to Use a Star Schema - A Demonstration Data Warehouse Schema- Industry Example
			Updating Table	Introduction Adding and Deleting Columns Changing Column Name and Data Type Creating Table from existing table Changing Constraints (Primary key) Changing Constraints (Foreign key) String Manipulation Date Manipulation
			Window Functions	Introduction Introduction to Windowing Functions Frames Named Windows Window Functions' Restrictions
			User Defined Functions and Stored Procedures	Introduction Introduction to User defined Functions User defined functions (Application) Introduction to Stored Procedures Stored Procedures (Application)
			Query Optimisation	Introduction Optimisation in Select Clause Optimisation in Where Clause Optimisation in Group by and Order by Optimisation in Joins Optimisation in Window Function
	Assignment SQL	*Apply the basics of investing and your knowledge of Data Science to determine when to buy and sell a stock.*	*Assignment - Stock Market Analysis*	Problem Introduction Data Set Grading Criteria Submission
	Data Analysis in Excel	*Taught by one of the most renowned data scientists in the country (S.Anand, CEO, Gramener), this module takes you from a beginner level Excel user to an almost professional user.*	*Introduction to Excel*	Introduction Understanding the Excel Interface Slicing and Dicing Data - Sort and Filter Report Making I: Basic Formatting Report Making II: Conditional Formatting Report Making III: Advanced Formatting Printing and Page Layout Passwords and Naming Files Ungraded Assignment
			Data Analysis in Excel - I	Introduction Delimited Files Discovering Shortcuts Introduction to Formulae Complex Functions Cell Referencing and Text Functions Logical Formulae Anand's Anecdotes Creating and Formatting Charts Types of Charts Anecdotes - II
			Data Analysis in Excel - II	Introduction Creating a Pivot Table Analysing Data in a Pivot Table Filtering Data in a Pivot Table Anand's Anecdotes - Pivot Tables VLOOKUP - Linking Data from multiple files & tables Anand's Anecdotes - VLOOKUP Common Errors in Excel Anand's Anecdotes Ungraded Assignment
	Visualisation using Tableau	*Learn an important and widely used tool for Data Analysts - Tableau.*	*Data Exploration in Tableau*	Introduction Data Formats and Tableau Interface Connecting to the Data Data Preparation in Tableau Hierarchies and Drill Down
			Visualising and Analysing Data in Tableau	Introduction Bar Charts Scatter Plots and Pie Charts Tree Maps Dual Axes Charts
			Visualising and Analysing Data with Tableau - II	Introduction Histograms Box Plots Area Maps Calculations in Tableau Dashboard and Stories
	Analytics Problem Solving	*This module covers concepts of the CRISP - DM framework for business problem solving.*	*The CRISP-DM Framework - Business and Data Understanding*	Introduction Define the Business Problem - Business Understanding Owning an IPL Team - Business Understanding Understanding Raw Data Preparing Data for Analysis
			CRISP-DM Framework - Data Preparation, Modelling, Evaluation and Deployment	The Heart of Data Analysis: Modelling Model Evaluation and Deployment
	Investment case study	*Apply the basics of investing and your knowledge of Data Science to determine when to buy and sell a stock*	*Investment Case Group Project*	Objectives Downloads Checkpoints - Part 1 Checkpoints - Part 2 Evaluation Rubric Final Submission

upGrad | iiit-b

PG Diploma in Data Science

(Program Curriculum)

For Prep Sessions + Batch Start Dates:
Please refer to upgrad.com

Note: This curriculum is subject to change based on inputs from IIITB and Industry

COURSE	MODULE NAME	DESCRIPTION	SESSION	SEGMENT
STATISTICS AND EXPLORATORY DATA ANALYSIS	**Inferential Statistics**	*Let us learn how to use a random sample data to describe and make inference about the population.*	*Basics of Probability*	Introduction: Inferential Statistics
				Introduction: Basics of Probability
				Random Variables
				Probability Distributions - I
				Probability Distributions - II
				Expected Value - I
				Expected Value - II
				Practice Questions
			Discrete Probability Distributions	Introduction: Discrete Probability Distributions
				Probability Without Experiment - I
				Probability Without Experiment - II
				Binomial Distribution
				Binomial Distribution (Examples)
				Cumulative Probability
				Practice Questions
			Continuous Probability Distributions	Introduction: Continuous Probability Distributions
				Probability Density Functions - I
				Probability Density Functions - II
				Normal Distribution
				Standard Normal Distribution
				Practice Questions
			Central Limit Theorem	Introduction: Central Limit Theorem
				Samples
				Sampling Distributions
				Properties of Sampling Distributions
				Sampling Distributions
				Central Limit Theorem
				Practice Questions - Part I
				Estimating Mean Using CLT
				Confidence Interval - Example
				Practice Questions - Part II
			Inferential Statistics - Practice Session	Introduction to Module
				Basics of Probability
				Joint Probability and Conditional Probability
				Bayes' Theorem
				Assessments I
				Standardized Normal Distribution and Z- Score
				Assessments II
				Introduction to Sampling Methods
				Sampling and Estimation
				Assessments III
	Hypothesis testing	*You'll learn how to test whether your assumptions about population data are true or not using the sample data.*	*Concepts of Hypothesis Testing - I*	Introduction
				Understanding Hypothesis Testing
				Null and Alternate Hypotheses
				Making a Decision
				Critical Value Method
				Critical Value Method - Examples
			Concepts of Hypothesis Testing - II	Introduction
				p-value Method
				p-value Method - Examples
				Types of Errors
			Industry Demonstration of Hypothesis Testing	Introduction
				T Distribution
				Two-Sample Mean Test
				Two-Sample Proportion Test
				A/B Testing Demonstration
				Industry Relevance
				Hypothesis testing in Python
			Hypothesis Testing - Additional Resources	Introduction
				Z-test
				T-Test
				Chi-Square Test
				P-Value Approach
				F-Test
	Assignment: Statistics and Hypothesis testing	*Assignment based on the concepts learnt in Inferential Statistics and Hypothesis Testing*	*Assignment*	General Guidelines
				Problem Statement
				Assignment Rubrics
				Submission
	Exploratory Data Analysis	*In this module, we shall learn a set of techniques to display data in such a way that interesting features will become apparent*	*Data Sourcing*	Introduction to EDA
				Introduction
				Public and Private Data
				Private Data
				Public Data
				Public Data Exercise
			Data Cleaning	Introduction
				Fixing Rows and Columns
				Missing Values
				Standardising Values
				Invalid Values
				Filtering Data
			Univariate Analysis	Introduction
				Data Description
				Unordered Categorical Variables - Univariate Analysis
				Ordered Categorical Variables - Univariate Analysis
				Quantitative Variables - Univariate Analysis
				Quantitative Variables - Summary Metrics
			Segmented Univariate	Introduction
				Introduction to Segmented Univariate Analysis
				Basis of Segmentation
				Quick way of Segmentation
				Comparison of Averages
				Comparison of Other Metrics
			Bivariate Analysis	Introduction
				Bivariate Analysis on Continuous Variables
				Business Problems Involving Correlation
				Practice Questions
				Bivariate Analysis on categorical variables
			Derived Metrics	Introduction
				What are Derived Metrics?
				Types of Derived Metrics: Type Driven Metrics
				Types of Derived Metrics: Business Driven Metrics
				Practice Questions
				Types of Derived Metrics: Data Driven Metrics
	Data Visualisation in Python	*Visualise distributions and summary statistics of data using Python's visualisation libraries matplotlib and seaborn.*	*Introduction to Data Visualisation*	Course Overview
				Introduction: Data Visualisation
				Visualisations - Some Examples
				Visualisations - The World of Imagery
				Understanding Basic Chart Types I
				Understanding Basic Chart Types II
			Basics of Visualisation	Introduction
				Data Visualisation Toolkit
				Components of a Plot
				Sub-Plots
				Functionalities of Plots
			Plotting Data Distributions	Introduction
				Univariate Distributions
				Univariate Distributions - Rug Plots
				Bivariate Distributions
				Bivariate Distributions - Plotting Pairwise Relationships
			Plotting Categorical and Time-Series Data	Introduction
				Plotting Distributions Across Categories
				Plotting Aggregate Values Across Categories
				Time Series Data
	Assignment: Uber Suppy-Demand Gap	*An assignment to study, visualise and solve uber supply-demand gap problem*	*Uber Supply-Demand Gap*	Problem Statement
				Evaluation Rubric
				Submission
	EDA case study	*Use the concepts of EDA to decipher which types of customers default on a loan*	*Gramener Case Study*	Problem Statement
				Evaluation Rubric
				Final Submission
	Additional resources	*Here, you will find all the additional content for the course as and when they are added to this module*	*Course Wrap - EDA and Statistics*	Course Wrap - EDA and Statistics
			Basics of Probability	Pre-Reads
				Optional Questions
			Discrete Probability Distributions	Pre-Reads
				Optional Questions
			Exploratory Data Analysis	Power Law
				Recommended Additional Content
				Election Data : Case Study

upGrad | iiit-b

PG Diploma in Data Science

(Program Curriculum)

For Prep Sessions + Batch Start Dates:
Please refer to upgrad.com

Note: This curriculum is subject to change based on inputs from IIITB and Industry

COURSE	MODULE NAME	DESCRIPTION	SESSION	SEGMENT
MACHINE LEARNING I	**Linear Regression**	*Regression helps us to determine the strength of the relationship between one dependent variable and a series of other changing variables.*	*Simple Linear Regression*	Introduction Introduction to Machine Learning Regression Line Best Fit Line Strength of Simple Linear Regression Simple Linear Regression in Python Coding Practice - Simple Linear Regression
			Multiple Linear Regression	Introduction Multiple Linear Regression Modelling in Python - I Modelling in Python - II Housing Case Study Derived Variables VIF - Variance Inflation Factor Housing Case Study Predictions Variable Selection Using RFE Assumptions of Linear Regression Feature Selection Coding Practice - Building a Multiple Linear Regression Model
			Industry relevance of linear regression	Introduction Linear Regression: Revision Prediction vs Projection Media Company Case Study Exploratory Data Analysis Model Building I Model Building II Model Building III Assessing the Model Interpreting the Results
	Linear Regression assignment	*Build a model to understand the factors car prices vary on and help a Chinese company enter the US car market.*	*Assignment- Linear Regression*	Problem Statement - Part I Problem Statement - Part II Evaluation Rubric Final Submission
	Logistic Regression	*In this module, you will study the theory of logistic regression, a machine learning technique for binary classification*	*Univariate Logistic Regression*	Module Introduction: Logistic Regression Introduction: Univariate Logistic Regression Binary Classification Sigmoid Curve Finding the Best Fit Sigmoid Curve - I Finding the Best Fit Sigmoid Curve - II Odds and Log Odds
			Multivariate Logistic Regression - Model Building	Introduction Making Predictions Model Building - Coding Exercise Model Evaluation Sensitivity and Specificity-I Sensitivity and Specificity - II Model Evaluation Metrics - Comprehension Model Evaluation Metrics - Coding Exercise Gain and Lift Charts KS Statistic
			Logistic Regression - Industry Applications - Part I	Introduction Getting familiar with Logistic Regression Nuances of Logistic Regression - Sample Selection Nuances of Logistic Regression - Segmentation Nuances of Logistic Regression - Variable Transformation-I Nuances of Logistic Regression - Variable Transformation-II Nuances of Logistic Regression - Variable Transformation-III Nuances of Logistic Regression - Variable Transformation-IV (Optional)
			Logistic Regression: Industry Applications - Part II	Introduction Commonly Faced Challenges in Implementation of Logistic Regression Model Evaluation (A Second Look) Model Validation and Importance of Stability Tracking of Model Performance Over Time
	Unsupervised learning: Clustering	*Here you will learn how to group elements into different clusters when you don't have any pre-defined labels to classify them.*	*Introduction to Clustering*	Introduction Understanding Clustering Practical Example of Clustering - Customer Segmentation
			K Means Clustering	Introduction Steps of the Algorithm K Means Algorithm K Means as Coordinate Descent K Means++ Algorithm Visualising the K Means Algorithm Practical Consideration in K Means Algorithm Cluster Tendency
			Executing K Means in Python	Introduction Data Preparation Making the Clusters Let's Have Some Fun Other Behavioural Segmentation Types
			Hierarchical Clustering	Introduction Hierarchical Clustering Algorithm Interpreting the Dendrogram Types of Linkages Cutting the Dendrogram & Analyzing the Clusters Industry Insights Let's have some fun
			Other Forms of Clustering	Introduction K-Mode Clustering K-Mode in Python K-Prototype in Python DB Scan Clustering Practice Question Gaussian Mixture Model
	Unsupervised Learning: Principal Component Analysis	*This module will cover the concepts of PCA, which is an unsupervised machine learning technique mainly used in dimensionality reduction. It will also cover practical applications of PCA in Python.*	*Principal Component Analysis*	Introduction The Why And What of PCA Building Blocks of PCA Illustration - Finding Principal Components Comprehension - Calculating the Principal Components Singular Value Decomposition SVD Example - Image Compression Practice Questions
			PCA in Python	Introduction PCA: Python Implementation Practical Considerations and Alternatives Optional Assignment (MNIST Dataset) Comprehension: PCA, SVD and Eigenvectors
	HR analytic case study	*Use your skills to predict which employee is going to leave the company in the near future.*	*HR Analytics Case Study*	Problem Statement Evaluation Rubric Submission
	Support Vector Machine (Optional)	*Learn the fundamentals of SVMs and use them to detect spam emails, recognise alphabets and more!*	*SVM - Maximal Margin Classifier*	Introduction Introduction to SVM Concept of a Hyperplane in 2D Practice Questions Concept of a Hyperplane in 3D Maximal Margin Classifier
			SVM - Soft Margin Classifier	Introduction The Soft Margin Classifier The Slack Variable Comprehension-1: Notion of Slack Variables Cost of Misclassification SVM R-Lab
			Kernels	Introduction Introduction to Kernels Mapping Nonlinear Data to Linear Data Feature Transformation The Kernel Trick R Lab - Kernels Shiny App - Types of kernels Choosing a Kernel Function Letter Recognition Using SVM

upGrad | IIIT-B

PG Diploma in Data Science

(Program Curriculum)

For Prep Sessions + Batch Start Dates:
Please refer to upgrad.com

Note: This curriculum is subject to change based on inputs from IIITB and Industry

COURSE	MODULE NAME	DESCRIPTION	SESSION	SEGMENT
MACHINE LEARNING II	**Tree Model**	*Tree models represent the way we make decisions. Learn how decisions are made in this powerful classification algorithm.*	*Introduction to Decision Trees*	Introduction
				Introduction to Decision Trees
				Interpreting a Decision Tree
				Comprehension - Decision Tree Classification in Python
				Regression with Decision Trees
			Algorithms for Decision Tree Construction	Introduction
				Concept of Homogeneity
				Gini Index
				Entropy and Information Gain
				Comprehension - Information Gain
				Splitting by R-squared
			Truncation and Pruning	Introduction
				Advantages and Disadvantages
				Tree Truncation
				Tree Pruning
				Building Decision Trees in Python
				Choosing Tree Hyperparameters in Python
				Coding Practice Questions
				Comprehension - Hyperparameters
			Random Forests	Introduction
				Ensembles
				Comprehension - Ensembles
				Creating a Random Forest
				Comprehension - OOB (Out-of-Bag) Error
				Comprehension - Time Taken to Build a Random Forest
				Random Forests Lab
				Coding Practice Questions
	Boosting	*This module will cover the concepts of boosting and different boosting algorithms- Adaboost, GBM and XGBoost.*	*Introduction to Boosting and AdaBoost*	Introduction
				Introduction to Boosting
				Weak Learners
				AdaBoost Algorithm
				AdaBoost Distribution and Parameter Calculation
				AdaBoost Lab
			Gradient Boosting	Introduction
				Understanding Gradient Boosting
				Gradient in Gradient Boosting
				Gradient Boosting Algorithm
				XGBoost
				Kaggle Practice Exercise
	Time Series (Optional)	*In this module, you will learn how to analyse and forecast a series that varies with time.*	*Intro to Time Series*	Time Series VS Regression
				Components of Time Series
			Working with Stationary Time Series	Understanding Stationarity
				Understanding White Noise
				Acf & Pacf Plots
				Ar & Ma Modelling
				Arma Modelling
				Model Evaluation
			End-to-end Analysis	Time Series Differencing
				Differencing VS Classical Decomposition
				Additive VS Multiplicative Model
				Time Series Smoothing
				Making Time Series Forecast
	Neural Networks (Optional)	*Inspired by the most sophisticated machine in the world - the human brain, NNs help machines learn.*	*Structure of Neural Networks*	Inspiration from Human Brain
				Working of A Neuron
				Hyper Parameters of Neural Networks
				Simplifying Neural Networks
				Specifying the Hyperparameters
				Activation Function
				Building a Sample Network on MNIST Data
			Information Flow in Neural Networks	Layers in Neural Networks
				Information Flow in Neural Networks
				Information Flow - Image Recognition
			Training a Neural Network	What does Training a Network Mean?
				Complexity of the Cost Function
				Updating the Weight & Biases
				Updating the Weights & Biases
			Training in Batches	Stochastic Gradient Descent
				Exploration & Exploitation
			Recurrent Neural Networks	Dealing with Sequential Data
				Regularisation in Neural Networks
	Model Selection	*You are preparing for a competitive exam. Should you learn some tricks for it or focus on the fundamentals? Model Selection has the answer*	*Principles of Model Selection*	Introduction
				Introduction to Model Selection
				Model and Learning Algorithm
				Simplicity, Complexity and Overfitting
				Bias-Variance Tradeoff
				Comprehension - Bias Variance Tradeoff
				Regularization
			Model Evaluation	Introduction
				Regularization and Hyperparameters
				Model Evaluation and Cross Validation
				Model Evaluation: Python Demonstration-I
				Model Evaluation: Python Demonstration-II
				Cross-Validation: Motivation
				Cross-Validation: Python Demonstration
				Cross-Validation: Hyperparameter Tuning
	Model Selection - Practical Considerations	*Given a business problem, how do you choose the best algorithm? Learn a few practical tips for doing this here*	*Model Selection - Best Practices*	Introduction
				Understanding the Business Problem
				Comprehension - Logistic Regression
				Comparing Different Machine Learning Models - I
				Comparing Different Machine Learning Models - II
				Pros and Cons of Different Machine Learning Models
				End-to-End Modelling - I
				CART and CHAID Trees
				Choosing between Trees and Random Forests - I
				Choosing between Trees and Random Forests - II
				End-to-End Modelling - II
	Advanced Regression	*This course takes a more advanced look at linear regression models.*	*Generalized Linear Regression*	Introduction
				Generalized Regression
				Generalized Regression Framework-1
				Generalized Regression Framework-2
				Systems of Linear Equations
				Generalized Regression Framework-3
				Generalized Regression in Python
			Regularized Regression	Introduction
				Regularized Regression
				Ridge and Lasso Regression - I
				Ridge and Lasso Regression - II
				Ridge and Lasso Regression in Python
				Model Selection Criteria-I
				Model Selection Criteria-II
				Feature Selection
				Comprehension - Model Selection Parameters
				Comprehension: Features' Subset Selection - Best Subset Selection
				Comprehension: Features' Subset Selection - Stepwise Selection
				Optional Assignment
	Telecom Churn Case Study	*Solve the most crucial business problem for a leading telecom operator in India and southeast Asia - predicting customer churn.*	*Telecom Churn - ML Group Case Study*	Problem Statement
				Evaluation Rubrics
				Submissions

upGrad | IIIT-B

PG Diploma in Data Science

(Program Curriculum)

For Prep Sessions + Batch Start Dates:
Please refer to upgrad.com

Note: This curriculum is subject to change based on inputs from IIITB and Industry

COURSE	MODULE NAME	DESCRIPTION	SESSION	SEGMENT
BIG DATA & SQL	**Introduction to Big Data**	*Understand the big data ecosystem and the various types of job roles in the industry.*	*Understanding Big Data*	Course Introduction Fundamentals of Big Data Identifying Big Data Conventional Data Processing Systems and Big Data
	Big Data Storage and Processing Framework - Hadoop	*Learn the basics of Hadoop and its architecture - a distributed computing platform.*	*Big Data Storage in Hadoop*	Introduction History of Hadoop Distributed Computing Hadoop Terminologies Master and Slave Hadoop Distributed File System Interaction Between Nodes in a Hadoop Cluster Advantages of Distributed File Systems Comprehension — Hadoop Distributed File System (HDFS)
			Big Data Processing In Hadoop	Introduction YARN - Yet Another Resource Negotiator MapReduce Comprehension - MapReduce Hadoop Vendors Hadoop Ecosystem Comprehension -Yet Another Resource Negotiator (YARN)
	Big Data Ingestion and Processing	*In big data ingestion and processing, learn to use various tools for getting and processing data.*	*Introduction to Apache Sqoop*	Introduction Data Ingestion with Apache Sqoop Advantages and Industry Use Cases of Sqoop How Sqoop Performs Import Comprehension: How Sqoop Import Works Creating an RDS Migrating Databases to the RDS Running Sqoop in AWS Adding a MySQL Connector Sqoop Commands: Listing Databases and Tables Sqoop Commands: Import and Import-All-Tables Sqoop Commands: Job and Eval
			Introduction to Apache Hive	Introduction Introduction to Apache Hive Key Features of Apache Hive Use Cases of Apache Hive The Hive Metastore Hive Data Models Creating Tables in Hive Understanding and Analysing the Data Stored in Hive Tables Solution - Movies Graded Questions
			Hive Data Models - Partitions and Buckets	Introduction Partitions Creating and Querying Partitioned Tables Buckets Comprehension: Data Models (Graded Assessment)
			File Formats in Apache Hive	Introduction File Formats in Apache Hive ORC and Compression Algorithms
			Advanced Data Analysis in Hive	Introduction EDA and UDFs in Hive Advanced Data Analysis using Hive Basic Text Analysis using Hive Handling Complex Data Types using Hive
	Big Data Processing using Apache Spark	*Learn Apache Spark, the newest big data framework with unprecedented performance and ease of use.*	*Concepts and Fundamentals of Spark*	Introduction Overview of Spark Spark vs MapReduce Resilient Distributed Datasets (RDDs) In-memory Processing RDD Operations Programming & Debugging in PySpark
			Working with Spark	Introduction: Setting Up Schema-on-Read v/s Schema-on-Write Comparing Spark With Hive Analysis with Spark - I: Reading & Summarising Data Analysis with Spark - II: Plotting Data Analysis with Spark - III: Filtering & Grouping Analysis with Spark - IV: Model-building Practice Analysis: Airlines Data MLlib - I: An Overview MLlib - II: Preparation for Model Building MLlib - III: Building ML models PySpark: An Alternative Library to PySpark Solution to PySpark Practice Questions Hive LLAP
	NYC Parking Case Study: Apache Spark	*Apply machine learning algorithms to Big Data using Spark*	*NYC Parking Tickets: An Exploratory Analysis*	Problem Statement Rubric Submission

upGrad | IIIT-B

PG Diploma in Data Science

(Program Curriculum)

For Prep Sessions + Batch Start Dates:
Please refer to upgrad.com

Note: This curriculum is subject to change based on inputs from IIITB and Industry

COURSE	MODULE NAME	DESCRIPTION	SESSION	SEGMENT
ELECTIVE- NATURAL LANGUAGE PROCESSING	**Lexical Processing**	***Get started with NLP by knowing about all the essential text preprocessing and text cleaning techniques.***	*Introduction to NLP*	NLP: Areas of Application
				Understanding Text
				Text Encoding
				Regular expressions: Quantifiers - I
				Regular Expressions: Quantifiers - II
				Comprehension: Regular Expressions
				Regular Expressions: Anchors and Wildcard
				Regular Expressions: Characters Sets
				Greedy versus Non-greedy Search
				Commonly Used RE Functions
				Regular Expressions: Grouping
				Regular Expressions: Use Cases
			Basic Lexical Processing	Word Frequencies and Stop Words
				Tokenisation
				Bag-of-Words Representation
				Stemming and Lemmatization
				Final Bag-of-Words Representation
				TF-IDF Representation
				Building a Spam Detector - I
				Building a Spam Detector - II
			Advanced Lexical Processing	Canonicalisation
				Phonetic Hashing
				Edit Distance
				Spell Corrector - I
				Spell Corrector - II
				Pointwise Mutual Information - I
				Pointwise Mutual Information - II
	Syntactic Processing	***Learn algorithms to parse grammar of sentences - HMMs, CFGs, PCFGs and build a smart flight-booking NLU system using techniques such as NER.***	*Introduction to Syntactic Processing*	The What and Why of Syntactic Processing
				Parsing
				Parts-of-Speech
				Different Approaches to POS Tagging
				Lexicon and Rule-based POS Tagging
				Stochastic Parsing
				The Viterbi Heuristic
				Markov Chain and HMM
				Explanation Problem
				Learning HMM Model Parameters
				HMM and the Viterbi Algorithm: Pseudocode
				HMM & the Viterbi Algorithm: Python Implementation
				Deep Learning Based POS Taggers
			Parsing	Why Shallow Parsing is Not Sufficient
				Constituency Grammars
				Top-Down Parsing
				Bottom-Up Parsing
				Probabilistic CFG
				Chomsky Normal Form
				Dependency Parsing
			Information Extraction	Understanding the ATIS data
				Information Extraction
				POS Tagging
				Rule-Based Models
				Probabilistic Models for Entity Recognition
				Naive Bayes Classifier for NER
				Decision Tree Classifiers for NER
				HMM and IOB labelling
				CRFs - Another Probabilistic Approach
			Conditional Random Fields	CRF Model Architecture - I
				CRF Model Architecture - II
				Training a CRF model
				Predicting using CRF
				Python Implementation of CRF
	Syntactic Syntactic Processing -Assignment Processing	***POS tagging is a crucial part of Syntactic Analysis. Build a POS tagger using a CRF classifier and by modifying Viterbi***	*Assignment - Syntactic Analysis*	Problem Statement
				Evaluation Rubric
				Final Submission
	Semantic Processing	***Extract meaning from the text***	*Introduction to Semantic Processing*	Concepts and Terms
				Entity and Entity Types
				Arity and Reification
				Schema
				Semantic Associations
				Databases - WordNet and ConceptNet
				Word Sense Disambiguation - Naive Bayes
				Word Sense Disambiguation - Lesk Algorithm
				Lesk Algorithm Implementation
			Distributional Semantics	Occurrence Matrix
				Co-occurrence Matrix
				Word Vectors
				Word Embeddings
				Latent Semantic Analysis (LSA)
				Comprehension - Latent Semantic Analysis
				Skipgram Model
				Comprehension - Word2Vec
				Generate Vectors using LSA
				Word2vec in Python - I
				Word2vec and GloVe in Python - II
				Word2vec and GloVe in Python - III
				Basics of Topic Modelling with ESA
				Introduction to Probabilistic Latent Semantics Analysis (PLSA)
			Topic Modelling	The Output of a Topic Model
				Defining a Topic
				Matrix Factorisation Based Topic Modelling
				Probabilistic Model
				Probabilistic Latent Semantic Analysis (PLSA)
				Expectation Maximization in PLSA
				Comprehension - Multinomial Distribution in Topic Modelling
				Latent Dirichlet Allocation (LDA)
				LDA - An extension of PLSA
				Use LDA to Generate a Corpus
				Parameter Estimation using Gibbs Sampling
				LDA in Python - I
				LDA in Python - II
				LDA in Python - III
			Social Media Opinion Mining - Semantic Processing Case Study	The Problem Statement
				Project Pipeline
				Python code - I
				Python code - II
	Building Chatbots With Rasa	***Learn the fundamentals of building chatbots using open source chatbot building framework -Rasa***	*Building Chatbots With Rasa*	Building Chatbots with Rasa
				Installation Guide - Rasa
				Natural Language Understanding (NLU)
				Training the NLU Model
				Dialogue-Flow Management
				Creating Conversational Stories & Defining Actions
				Training the Dialogue Management Model
				Interactive Learning
				Chatbot Deployment
				ML and AI in Business
			NLP Course Project - Building a Chatbot	Problem Statement
				Evaluation Rubric
				Final Submission

upGrad | iiit-b

PG Diploma in Data Science

(Program Curriculum)

For Prep Sessions + Batch Start Dates:
Please refer to upgrad.com

Note: This curriculum is subject to change based on inputs from IIITB and Industry

COURSE	MODULE NAME	DESCRIPTION	SESSION	SEGMENT
ELECTIVE- DEEP LEARNING AND NEURAL NETWORKS	**Introduction to Neural Networks**	*In this module, you'll be introduced to the basics of Neural Networks and various concepts related to Deep Neural Network which will be used in the future modules.*	*Structure of Neural Networks*	Neural Networks - Inspiration from the Human Brain
				Introduction to Perceptron
				Binary Classification using Perceptron
				Perceptrons - Training
				Multiclass Classification using Perceptrons
				Working of a Neuron
				Inputs and Outputs of a Neural Network I
				Inputs and Outputs of a Neural Network II
				Assumptions made to Simplify Neural Networks
				Parameters and Hyperparameters of Neural Networks
				Activation Functions
			Feed Forward in Neural Networks	Flow of Information in Neural Networks - Between 2 Layers
				Information Flow - Image Recognition
				Comprehension - Count of Pixels
				Learning the Dimensions Weight Matrices
				Feedforward Algorithm
				Vectorized Feedforward Implementation
				Understanding Vectorized Feedforward Implementation
			Backpropagation in Neural Networks	What Does Training a Network Mean?
				Complexity of the Loss Function
				Comprehension - Training a Neural Network
				Updating the Weights and Biases - I
				Updating the Weights and Biases - II
				Updating the Weights and Biases - III
				Sigmoid Backpropagation
				Updating the Weights and Biases - IV
				Updating the Weights and Biases - V
				Updating the Weights and Biases - VI
				Batch in Backpropagation
				Training in Batches
			Modifications to Neural Networks	Regularization
				Dropouts
				Batch Normalization
				Introduction to Keras
			Hyperparameter Tuning in Neural Networks	Loss Function I
				Loss Function II
				Minibatch Gradient Descent
				Gradient Descent I
				Gradient Descent II
				Gradient Descent III
				Gradient Descent IV
				Momentum based methods I
				Momentum based methods II
				Momentum based methods III
				Dropouts -The Bayesian Approach
				Vanishing and Exploding Gradients
				Initializations
	Neural Networks - Assignment	*Implementing multiclass classification on MNIST dataset using raw NN model in Numpy.*	*Introduction to Neural Networks- Assignment*	Problem Statement
				Evaluation Rubric
				Final Submission
	Convolutional Neural Networks	*Learn how to solve state-of-the-art computer vision problems using CNNs.*	*Introduction to Convolutional Neural Networks*	A Specialised Architecture for Visual Data
				Applications of CNNs
				Understanding the Visual System of Mammals - I
				Understanding the Visual System of Mammals -II
				Introduction to CNNs
				Reading Digital Images
				Video Analysis
				Understanding Convolutions - I
				Understanding Convolutions - II
				Stride and Padding
				Important Formulas
				Weights of a CNN
				Feature Maps
				Pooling
				Putting the Components Together
			Building CNNs with Python and Keras	Building CNNs in Keras - MNIST
				Comprehension - VGG16 Architecture
				CIFAR-10 Classification with Python - I
				CIFAR-10 Classification with Python - II
				CIFAR-10 Classification with Python - III
			CNN Architectures and Transfer Learning	Overview of CNN Architectures
				AlexNet and VGGNet
				GoogleNet
				Residual Net
				Introduction to Transfer Learning
				Use Cases of Transfer Learning
				Transfer Learning With Pre-Trained CNNs
				Practical Implementation of Transfer Learning
				Transfer Learning in Python
				An Analysis of Deep Learning Models - I
				An Analysis of Deep Learning Models - II
			Style Transfer and Object Detection	Introduction to Style Transfer
				Style Loss and the Gram Matrix
				Loss Function
				Style Transfer Notebook
				Object Detection - I
				Object Detection - II
	Convolutional Neural Networks - Industry Applications	*Learn about how CNNs are used in industry*	*Industry Demo: Using CNNs with Flowers Images*	Examining the Flowers Dataset
				Data Preprocessing: Shape, Size and Form
				Data Preprocessing: Normalisation
				Data Preprocessing: Augmentation
				Data Preprocessing: Practice Exercise Solutions
				ResNet: Original Architecture and Improvements
				Building the Network
				Ablation Experiments
				Hyperparameter Tuning
				Training and Evaluating the Model
			Industry Demo: Using CNNs with X-ray Images	Examining X-ray images
				CXR Data Preprocessing - Augmentation
				CXR Data Preprocessing - Normalisation
				CXR: Network Building
				CXR: Final Run
	Recurrent Neural Networks	*Learn how to use neural networks on sequence problems using recurrent neural networks.*	*What Makes a Neural Network Recurrent?*	What are Sequences?
				What Makes the Network Recurrent
				Architecture of an RNN
				Feeding Sequences to RNNs
				Comprehension: RNN Architecture
				Types of RNNs - I
				Training RNNs
				Types of RNNs - II
				Vanishing and Exploding Gradients in RNNs
			Variants of RNNs	Bidirectional RNNs
				Long, Short-term Memory Networks
				Characteristics of an LSTM Cell
				Structure of an LSTM Cell
				LSTM Network: Feedforward Equations
				GRUs and Other Variants
			Building RNNs in Python	POS Tagging Using RNN - I
				POS Tagging Using RNN - II
				POS Tagging Using RNN -III
				POS Tagging Using RNN -IV
				POS Tagging Using RNN -V
				Generating C Code - I
				Generating C Code - II
				Generating C Code - III
				RNNs in Python
	Neural Networks Project - Gesture Recognition	*In this module, you'll experiment and create a model that identifies the gestures with considerable accuracy.*	*Deep Learning Course Project - Gesture Recognition*	Problem Statement
				Two Architectures: 3D Convs and CNN-RNN Stack
				Understanding Generators
				Starter Code Walkthrough
				Evaluation Rubric
				Final Submission

upGrad | iiit-b

PG Diploma in Data Science

(Program Curriculum)

For Prep Sessions + Batch Start Dates:
Please refer to upgrad.com

Note: This curriculum is subject to change based on inputs from IIITB and Industry

COURSE	MODULE NAME	DESCRIPTION	SESSION	SEGMENT
ELECTIVE- HEALTHCARE	**Understanding the Healthcare Domain**	*With all the necessary DA knowledge, it is time to get into the domain details. Learn about the healthcare landscape in the US.*	*What to expect*	What to expect from Healthcare Domain Elective
			Introduction to the Healthcare Space	Introduction
				Understanding the Healthcare Market
				Stakeholders of the Primary Healthcare Ecosystem: Process
				Stakeholders of the Primary Healthcare Ecosystem: Drivers and Metrics
				Stakeholders of the Secondary Healthcare Ecosystem: Process
				Stakeholders of the Secondary Healthcare Ecosystem: Drivers and Metrics
				Other Stakeholders of the Healthcare Ecosystem
	Provider Analytics	*In this module, you will explore the different analytics opportunities that exist in the healthcare provider space.*	*Provider Analytics*	Introduction
				Analytics Related to Patient-Physician Interactions
				Clinical Decision Support Systems
				Analytics Related to Patient-Hospital Interactions
				Management of Patient Traffic - I
				Management of Patient Traffic - II (Comprehension)
				Management of Patient Traffic - III
				Hospital Performance Analysis - I
				Hospital Performance Analysis - II (Comprehension)
				Hospital Performance Analysis - III
				Hospital Compare
	Payer Analytics	*In this module, you will explore the different analytics opportunities that exist in the healthcare payer space.*	*Getting Familiar with the US Payer Market*	Introduction
				Payers in the US
				Types of Health Insurance
				Types of Insurance Plans
				Benefits
				Analytics Opportunities in Benefits
				Coordination of Benefits
				Provider Management - I
				Provider Management - II
				Pay for Performance (P4P)
				Analytics Opportunities in Provider Management
			Claims and Care Management	Introduction
				Life Cycle of a Health Insurance Claim
				Healthcare Coding
				Claims Adjudication
				Analytics Opportunities in Claims Management
				Analytics to Detect Fraudulent Claims
				Care Management
				Care Management Framework
				Risk Stratification
				Evaluating a Care Management Program
				Accountable Care Organisations (ACOs)
				Analytics Opportunities in Care Management
	Assignment - Payer Analytics	*Stratify patients according to the risk of cost they pose to the healthcare payer*	*Assignment - Risk Stratification*	Problem Statement
				Submission
	Analytics in the Pharmaceutical Industries	*Learn how pharmaceutical companies harness the power of data analytics.*	*Drug Development and Sales Analytics*	Introduction
				Pharmaceutical Market Overview
				Drug Development Life Cycle
				Areas of Analytics in Pharma
				Pharmaceutical-Selling Process
				Field Activity
				Analytics in Sales
				Sales Data
				Customer Segmentation
			Marketing Analytics	Introduction
				Structure of a Marketing Organisation
				Multichannel Marketing (MCM) Management
				Patient Journey Analytics
				Analytics Opportunities in Commercial Operations
				Market Forecasting
	Course Wrap for Healthcare	*Get a brief overview of how all that you have studied in the healthcare domain, finds application in the real world.*	*Course Wrap*	Healthcare Course Wrap by Prof. RC
				Interview tips by Rohit
	Capstone Project	*Decipher the CMS hospital star rating system using supervised and unsupervised models.*	*Capstone- Healthcare*	Problem Statement
				Mid Submission
				Final Submission

upGrad | iiit-b ज्ञानमुत्तमम्

PG Diploma in Data Science

(Program Curriculum)

For Prep Sessions + Batch Start Dates:
Please refer to upgrad.com

Note: This curriculum is subject to change based on inputs from IIITB and Industry

COURSE	MODULE NAME	DESCRIPTION	SESSION	SEGMENT
ELECTIVE- E-COMMERCE	**Introduction to E-commerce**	***Get acquainted with the various applications of Data Analytics in E-commerce business***	*Introduction to ecommerce domain*	Introducing Analytics in E-Commerce
			Data Analytics in ecommerce	Introduction
				Business of ecommerce
				Inventory Management
				Marketing in ecommerce
				Improving User Experience
				Fraud Detection
				Shipment Delivery
				Customer Feedback
	Recommendation systems	***Learn about the algorithms that power the recommendation engines of the E-commerce sites***	*Recommendation Systems*	Introduction
				Understanding Recommendation Systems
				Content Based Filtering
				User Based Collaborative Filtering
				Item Based Collaborative Filtering
				Issues in Recommendation Systems
				Recommender System in Python
	Assignment - Recommendation systems	***Build a recommendation engine based on Beer preferences of users.***	*Assignment - Recommendation System*	Problem Statement
				Submission
	Price Optimization	***Learn how prices are dynamically optimised on an e-commerce platform***	*Price Markup & Markdown*	Introduction
				Understanding Price Markup & Markdown
				Why are Markdowns done?
				Why are Markups done?
				Effect of Price Markup or Markdown
			The Four-Force Model	Introduction
				The Four Forces of Price Optimisation
				Demand elasticity
				Competitive benchmark
				Internal economics
				Category dynamics
				Goal of Price Optimisation
	Market Mix Modelling	***Learn how to optimise your marketing spends in order to maximise the ROI.***	*Factors that Impact Sales*	Introduction
				What is Market Mix Modelling (MMM)?
				How Does Advertising Impact Revenue?
				How Do Pricing & Promotions Impact Revenue?
				How Does Product Assortment Impact Revenue?
			Modelling the Impact of KPIs	Introduction
				Modelling the Advertising Effects - Part I
				Modelling the Advertising Effects - Part II (Optional)
				Modelling the Advertising Effects - Part III
				Creating AdStocks
				Modelling Different Pricing Effects
				Overview of KPIs
				Presenting the Results
				Other Topics
	A\B Testing (Optional)	***Understand the concept behind A/B test and also learn how to execute an A/B test in Optimizely***	*A/B Testing*	Introduction
				Understanding A/B testing
				Steps in A/B testing
				Setting up an A/B Test in Optimizely
	Course wrap for E-commerce	***Get a brief overview of how all that you have studied in the ecommerce domain, finds application in the real world.***	*Course Wrap*	Ecommerce Course Wrap by Ujjyaini
	Capstone Project	***Model the impact of different marketing levers on the sales figure of ElecKart.***	*Capstone- E-com*	Problem Statement
				Mid Submission
				Final Submission

upGrad | IIIT-B ज्ञानमुत्तमम्

PG Diploma in Data Science

(Program Curriculum)

For Prep Sessions + Batch Start Dates:
Please refer to upgrad.com

Note: This curriculum is subject to change based on inputs from IIITB and Industry

ELECTIVE - BFSI

COURSE	MODULE NAME	DESCRIPTION	SESSION	SEGMENT
	Introduction to Banking and Financial Services	***Learn how banks make money through various banking products and also understand the customer lifecycle.***	*What to Expect in BFS Domain Elective*	Introducing Analytics in Banking and Financial Services
			How Banks Make Money	Introduction
				Banking Products - Deposits & Lending
				Profitability of Credit Cards
				P&L of Banks and Financial Institutions
			Customer Lifecycle	Introduction
				Customer Lifecycle
				Customer Lifecycle - Acquisition Analytics
				Customer Lifecycle - Engagement Analytics
				Customer Lifecycle - Risk Analytics
	Acquisition Analytics	***Understand the component of Acquisition Strategies & practice hands-on exercise of Data analytics for acquiring the potential customers***	*Introduction to Acquisition Analytics*	Introduction
				Acquisition - Types of Datasets
				Components of Acquisition Strategy
				Market Segmentation
				Segment Prioritisation
				Channel Preferences
	Assignment- Acquisition Analytics	***Build a response model based on the clients, campaign and economic information provided by the portuguese bank.***	*Assignment - Acquisition Analytics*	Problem Statement
				Submission
	Engagement Analytics	***Now that you have learnt how to acquire customers, learn how to engage them and prevent their attrition***	*Engagement Strategies*	Introduction
				Engagement Analytics Framework
				Cross-Selling Strategies
				Types of Cross-Selling
				Cross-Selling Opportunities
				Customer Lifetime Value (CLV)
			Cross-Selling Lab	Introduction
				Cross-Selling - Business Objectives
				Cross-Selling Analysis
			Retention and Loyalty Management	Introduction
				Types of Attrition
				Attrition - Credit Card
				Interpreting a Credit Card Attrition Model
	Risk analytics	***Learn about the risk associated with customers who default on their loan or credit, and the analytics related to it.***	*Types of Risk Analytics*	Introduction to Risk Analytics
				Types of Risk Analytics
				Types of Credit Risk - Operational and Regulatory
			Operational Risk Analytics - Acquisition	Introduction to Operational Risk Analytics
				Operational Risk Analytics Framework - Customer Lifecycle
				Introduction to Acquisition Risk Analytics
				Metrics to Measure Acquisition Risk
				Implementing Acquisition Risk Models
				Validating Acquisition Risk Models
			Operational Risk Analytics - Existing Customer Management	Introduction
				Roll Rate Matrix
				Data and Models in ECM
				Validation of Behaviour Models
			Operational Risk Analytics - Collection & Recovery	Collection and Recovery Management
				Validation of Collection and Recovery Models
			Regulatory Risk Analytics	Introduction
				Regulatory Risk Analytics - A Brief Introduction (Optional)
	Course wrap for BFS	***Get a brief overview of customer lifecycle that you have studied in the BFS domain, finds its application in the real world.***	*Course Wrap*	BFS Course Wrap by Kalpana
	Capstone Project	***Help CredX identify the ideal applicants to provide credit cards to by building an application scorecard.***	*Capstone- BFS*	Problem Statement
				Mid Submission
				Final Submission

upGrad | IIIT-B ज्ञानमुत्तमम्

PG Diploma in Data Science

(Program Curriculum)

For Prep Sessions + Batch Start Dates:
Please refer to upgrad.com

Note: This curriculum is subject to change based on inputs from IIITB and Industry

COURSE	MODULE NAME	DESCRIPTION	SESSION	SEGMENT
CAPSTONE	**Introduction to Kaggle**	***An introduction to the world of Kaggle. How it can be used to enhance visibility.***	*Introduction to Kaggle*	Introduction to Kaggle Creating an account Datasets Kernels Competitions
	Feature Engineering	***Build general features to build a model for text analytics***	*Basic Feature Engineering*	Basic Feature Engineering
			Advanced Feature Engineering	Advanced Feature Engineering Model Building
	Capstone	***Solve a problem based on one of the competitions held on Kaggle or on an industry dataset as a final test of what you have learned so far.***	*Problem statement*	Problem Statement Evaluation Rubric Final Submission
			Solution	Solution