

OUTLINE FOR THE POST GRADUATE PROGRAM IN DATA SCIENCE & MACHINE LEARNING (PGPDM)

Module	Subject	Topics	Learning outcomes	Delivered by
Exploratory Data Analysis & Visualization	Data Analysis Framework	<ul style="list-style-type: none"> • Introduction to evidence based decision making • Business problems and data analysis • Data and Analytical thinking • Introduction to XLMiner 	<p>At the end of this module, a student will</p> <ul style="list-style-type: none"> * Gain an understanding of data mining and data analysis in general * Know how to apply a data analysis framework to business problems * Be able to assess data quality and sufficiency for proceeding with data analysis * Know how to be able to create representative samples for data analysis and when they are required * Perform exploratory data analysis * Learn how to visualize patterns and relationships in data, and create advanced dashboards for reporting and insights * Be able to use SPSS modeler for EDA and visualization * Be exposed to Watson analytics capabilities for EDA and visualization tasks 	UC
	Exploratory Data Analysis	<ul style="list-style-type: none"> • Assess data quality and data coverage • Exploratory Data Analysis (EDA) • Exploratory Data Analysis 		
	Data Collection and Curation	<ul style="list-style-type: none"> • Data curation: building data analysis universe • Sampling techniques • Creating modeling universe using Microsoft Excel and XLMiner 		
	Data Visualization	<ul style="list-style-type: none"> • Visualization • Visualizing Relationships • Spotting Differences • Tableau for Data Visualization 		
R for Data Science	Introduction to R	<ul style="list-style-type: none"> • An introduction to R • Importing and exporting data in R from multiple sources and file types • Data Storage and processing in R 	<ul style="list-style-type: none"> * Be able to do common SQL like tasks such as filtering rows, selecting columns, ordering data and doing group by tasks * Be able to reshape data from long to wide format and vice versa. * Be able to do simple string processing using standard R string functions * Be able to do univariate and bivariate analysis using standard R plotting functions * Be able to create highly customized visualizations using grammar of graphics visualization paradigm * Be able to use R for the entire data exploration and preparation process 	JA
	Data Manipulation and Visualization with R	<ul style="list-style-type: none"> • Data Manipulation with R • Advanced Data Manipulation with R • Data Visualization with R 		
	Data Pre-Processing with R	<ul style="list-style-type: none"> • Analytics Methodology • Pre-processing data for data science - data exploration and data preparation 		

Module	Subject	Topics	Learning outcomes	Delivered by
Data Analysis for Evidence Based Decision Making	Role of Uncertainty	<ul style="list-style-type: none"> • Role of uncertainty in decision making • Probability as measure of uncertainty • Probability distributions: normal, binomial, Poisson and exponential • Random variables, their main characteristics • Examples 	<p>At the end of this module, a student will</p> <ul style="list-style-type: none"> * Understand the role of statistics and statistical inference in data mining and analytics * Be able to identify multiple types of random variable outcome distributions and know how to calculate probabilities of outcomes based on distribution types * Know how to apply probability concepts to evaluate decision payoffs and make the appropriate choices * Learn about estimation as an important outcome of data mining, and know how to generate estimates as well as confidence intervals associated with the estimates * Learn how to test business hypothesis with a hypothesis testing framework, and how and when to apply different types of hypothesis tests * Build linear regression and ANOVA models, including data preparation tasks associated with model building 	UC
	Decision Making under Uncertainty	<ul style="list-style-type: none"> • Elements of decision analysis: payoff tables, decision criteria, expected monetary value, sensitivity analysis, decision trees, risk profiles • Bayes rule • Multistage decision problems and the value of information • Examples 		
	Statistical Inference	<ul style="list-style-type: none"> • Sampling and sampling distributions • Introduction to estimation • Confidence intervals estimation • Hypothesis testing • Examples 		
	Regression Analysis: Estimating Relationships	<ul style="list-style-type: none"> • Linear versus nonlinear relationships • Outliers • Unequal variance • Correlations: indicators of linear relationships • Simple linear regression • Examples 		
	Multiple Regression	<ul style="list-style-type: none"> • Interpretation of regression coefficients • Interpretation of standard error of estimate & R-Square • Modeling possibilities: dummy variables, nonlinear transformations • Statistical inference for linear model • Stepwise regression • Violations of linear model assumptions • Examples 		
	ANOVA and Experimental Design	<ul style="list-style-type: none"> • One-Way ANOVA • Using linear model to perform ANOVA • The multiple comparison problem • Two-Way ANOVA • Experimental design • Examples 		

Module	Subject	Topics	Learning outcomes	Delivered by
Data Mining	Creating Business Value using Data Mining for Actionable Insights	Introduction to Data Mining, and Data Visualization	<p>At the end of this module, a student will</p> <ul style="list-style-type: none"> * Thoroughly understand how data mining is performed on business data to generate actionable insights * Understand how to work with high dimension data, and when and how to apply Factor Analysis to reduce dimensions to a manageable number * Understand classification as an outcome of data mining, and the concepts and application of multiple classification techniques including k-NN, Naive Bayes, Logistic regression, and Decision Trees * Understand what is a data mining model, and how it can be used to make predictions * Understand how to assess models for accuracy, validity, and viability, and how to use ensemble methods to generate high levels of accuracy and validity 	UC
		Dimension Reduction via Factor Analysis using R		
		Prediction via K-NN, Naive Bayes, and Logistic Regression using R		
		Classification and Regression Trees using R		
		Model Evaluation, Model Validation, and Ensemble Modeling		
		Association Rules and Collaborative Filtering video		
Industry Applications of Advanced Analytics Models with Python	Implementing Predictive Analytics Models for Industry Use Cases	<ul style="list-style-type: none"> • Linear Regression Models in R, using Market Mix Case Study • Logistic Regression Models in R, using a Credit Score card case study • Time Series forecasting in R, using a Demand Forecasting Case Study 	<p>At the end of this module, a student will</p> <ul style="list-style-type: none"> * Be able to implement multiple predictive modeling and machine learning algorithms on business datasets * Understand how to choose appropriate algorithms for specific outcomes * Use R as a tool for data pre-processing, solution design, model building and model validation across multiple algorithms * Understand how solutions are built across multiple domains and functional areas spanning marketing, finance, telecom, operations, retail etc 	JA
	Implementing Machine Learning Models for Industry Use Cases	<ul style="list-style-type: none"> • Tree models in Python, with a Telecom Churn Case Study • Segmentation Models in Python, Retail Store Profiling Case Study • Ensemble methods in Python, with multiple case studies • SVM in Python, Handwritten Digit Recognition Case Study 		
Deep Learning Models	Advanced Machine Learning Models and Introduction to AI	<ul style="list-style-type: none"> • Neural networks and Deep Neural Networks • Tensor Flow • Using IBM Blue Mix and Watson platforms 	<ul style="list-style-type: none"> * Understand advances in Machine Learning models that enable multi-layer Deep Learning applications * Be exposed to the latest tools and techniques in Deep Learning via IBM platforms and open source platforms * Gain a general understanding of AI and application 	IBM

Module	Subject	Topics	Learning outcomes	Delivered by
Big Data Analytics with Spark	Introduction to Big Data	<ul style="list-style-type: none"> Understanding Big Data eco-system Implementing ML in Big Data 	<p>At the end of this module, a student will</p> <ul style="list-style-type: none"> * Understand what Big Data is and why it needs specialized technology to store, manage, process and analyze * Learn about the challenges associated with implementing data mining and ML algorithms on Big Data systems * Gain expert level understanding of Spark, a Big Data technology that allows implementation of ML algorithms on Big Data * Use SparkR and Spark MLLib to implement multiple ML algorithms in Spark 	JA
	ML in Spark	<ul style="list-style-type: none"> Introduction to Spark Spark Architecture SparkR MLlib in Spark and examples 		
Project Management in Analytics	Managing impactful analytics projects for internal clients	<ul style="list-style-type: none"> Aligning project deliverables with corporate objectives Finding and communicating with internal champions Fully leveraging internal culture and processes for project success Sharing results and driving change 	<p>At the end of this module, a student will</p> <ul style="list-style-type: none"> * Understand different techniques on how to set the project up for success * Learn communication skills that will smooth the way for project management * Know how to build a strong bond within organizations to ensure access to data, resources and partners needed to complete a project * Have a tool kit of change management techniques to drive positive impact in organizations by implementing project recommendations * Know how to make a build/buy decisions for analytic projects * Be able to scope, define, write a request for a proposal, negotiate, choose and on-board a third-party vendor * Learn how to tailor messaging to different stakeholders * Learn how to leverage successful project completion to advance their career 	UC
	Best in class project management to delight your external clients	<ul style="list-style-type: none"> Setting up client expectations Working with the client on obtaining data and access to the organization Client management tips and tricks Sharing project results to drive positive change 		
	Successful project outsourcing	<ul style="list-style-type: none"> Approaches to managing a blended project Project scoping and vendor selection process Vendor, project and internal management techniques Integration of internal and external outputs 		
Information to Insight	Advanced Visualization	<ul style="list-style-type: none"> Advanced visual analytics Advanced visualization topics Clustering and classification using Microsoft Excel and XLMiner 	<p>At the end of this module, a student will</p> <ul style="list-style-type: none"> * Be able to understand how data science results should be communicated to a business audience * Learn how to translate statistical and ML model results to business understandable results * Know how to identify actionable insights and generate business recommendations from data mining models * Use appropriate and effective presentation techniques to present results from data mining projects 	UC
	Storytelling with Data	<ul style="list-style-type: none"> Communicating data science results Effective presentation skills Using data visualizations for storytelling Delivering visual stories Tools for effective visual stories 		JA

Module	Subject	Topics	Learning outcomes	Delivered by
Career Management	Taking your career to the next level	<ul style="list-style-type: none"> • Planning your career • Resume writing • Interview preparation • Interview follow ups 	<p>At the end of this module, a student will</p> <ul style="list-style-type: none"> * Be able to plan their next career move based on overall career strategy * Know how to build a resume for a specific job/company/industry * Use effective interviewing techniques and avoid common pitfalls * Have a toolkit of effective tactics to seal getting an offer post-interview * Plan out their first 90 days on the new job to ensure successful start 	UC
Capstone Project			<p>At the end of the Capstone Project a student will</p> <ul style="list-style-type: none"> * Be able to demonstrate capability across all stages of an analytics project, including problem identification, solution design, implementation, and business outcome generation, using a live project and data shared by a client company * Use multiple tools as appropriate including R, SPSS, Spark, Excel to perform data mining and present results * Present an actionable and implemented analytics solution to the problem identified by the client project sponsor 	UC/JA