Making Data-Driven Decisions
Data Science and Big Data Analytics

Self-Paced Course

Price: ₹ 54,000/-  
Start Date: 3rd February, 2020  
Duration: 7 Weeks

CEUs: 1.8
MAKE BETTER DATA-DRIVEN BUSINESS DECISIONS

Could you be using your data more effectively?

90% of the world’s data has been created in just the past few years. Faced with overwhelming amounts of data, organizations are struggling to extract the powerful insights they need to make smarter business decisions. To help uncover the true value of your data, MIT Institute for Data, Systems, and Society (IDSS) created the online course Data Science and Big Data Analytics: Making Data-Driven Decisions for data scientist professionals looking to harness data in new and innovative ways.

Over the course of seven weeks, you will take your data analytics skills to the next level as you learn the theory and practice behind recommendation engines, regressions, network and graphical modeling, anomaly detection, hypothesis testing, machine learning, and big data analytics.

PROGRAM VITALS

Course Starts
3rd February, 2020
Self-Paced Course, 7 Weeks

Time Commitment
4-5 Hours Per Week

Program Fee
₹ 54,000/- + Applicable taxes

Debit and Credit cards accepted
Loan Facility Available

Visit website to pay
WHO SHOULD PARTICIPATE

- Professionals at any career stage, looking to turn large volumes of data into actionable insights.
- Past learners’ job roles have included: business intelligence analysts, management consultants, technical managers, business managers, data science managers.
- Data science enthusiasts and IT professionals.

Participants may include:
Technical managers | Business intelligence analysts | Management consultants
IT practitioners | Business managers | Data science managers | Data science enthusiasts

AFTER THIS COURSE, YOU WILL BE ABLE TO:

- Apply data science techniques to your organization's data management challenges
- Determine the difference between graphical models and network models.
- Convert datasets to models through predictive analytics.
- Deploy machine learning algorithms to improve business decision making.
- Master best practices for experiment design and hypothesis testing.
- Identify and avoid common pitfalls in big data analytics.
COURSE OUTLINE

Course materials blend the following pedagogical strategies to best achieve the learning objectives of the course and individual modules:

**Instructivism:** Teacher-centered learning where the instructors present relevant content (tutorial videos enhanced with animation and graphics). Students will test their knowledge through graded tests.

**Constructivism:** Learning by doing approach. We encourage learners to construct their own understanding through solving the mandatory and optional case studies and practicing.

**Social Constructivism:** Learning through social interactions and communication. You will be able to discuss with your peers in the discussion groups, and evaluate and get reviews from your peers through two compulsory case studies.

**Connectivism:** Connecting with others and extending your knowledge through communication. You will be able to expand and share your knowledge with others through the Discussion group, and course groups on Facebook, and LinkedIn.

---

**Module 1:**
- Making sense of unstructured data
  - Clustering
  - Spectral Clustering, Components and Embeddings
  - Case Studies

**Module 2:**
- Regression and Prediction
  - Classical Linear & nonlinear regression & extension
  - Modern Regression with High-Dimensional Data
  - The use of modern Regression for causal inference
  - Case Studies

**Module 3:**
- Classification, Hypothesis Testing and Anomaly Detection
  - Hypothesis Testing and Classification
  - Deep Learning
  - Case Studies

**Module 4:**
- Recommendation Systems
  - Recommendations and ranking
  - Collaborative filtering
  - Personalized recommendations
  - Case Studies
  - Wrap-up: Parting remarks and challenges
Module 5:
Networks and Graphical Models
- Introduction
- Networks
- Graphical Models
- Case Studies

Module 6:
Predictive Modeling for Temporal Data
- Introduction
- Prediction engineering
- Feature engineering
- Modeling and evaluating predictive models

About MIT xPRO

MIT xPRO's online learning programs leverage vetted content from world-renowned experts to make learning accessible anytime, anywhere. With the use of cutting-edge research in the neuroscience of learning, MIT xPRO programs are application-focused, helping professionals build their skills on the job.

Pearson, in collaboration with MIT xPRO, will distribute this course to learners in India. Payments will be processed by Pearson and all courseware delivered by MIT xPRO.

WHY MIT xPRO?

It's professional development– the MIT way.

MIT xPRO courses provide professional development opportunities to individuals, teams, and companies across the world. Leveraging the latest learning technologies, MIT xPRO courses and programs are designed to provide a high-quality education experience while accommodating your busy life. MIT xPRO learners are not only scientists, engineers, technicians, managers, and consultants– they are change agents. They take the initiative, push boundaries, and define the future.

EARN A CERTIFICATE OF COMPLETION AND CEUS

Participants who successfully complete the course and all assessments will receive a Certificate in Data Science from MIT xPRO and 1.8 Continuing Education Units (CEUs). This course does not carry MIT credits or grades, however, a 60% pass rate is required in order to receive the certificate. Course requirements include:

- Submission and peer-review of two case studies
- Passing grades on eight assessments
CASE STUDY OUTLINE

In this course, you won't just discover new strategies, tools, and insights— you'll put them to the test. Every course module features a selection of case studies and hands-on projects that help you apply your newfound knowledge to realistic business challenges.

COURSE REVIEWS

“Leveraging this knowledge will allow me to position myself as a hybrid analyst-data scientist, which greatly increases my value to the company.” - Ryan Michael Dickinson

“I really enjoyed the interactions/animations in the videos. These really helped with visualizing the concepts... I feel more equipped to understand what type of insights can be gleaned from a particular set of data, and can better communicate these asks to our data science team.” - Reza Dawood

“The course content was really amazing and gave me exact direction to head towards the Big Data topic.” - Prasad Sankpal

“It's very critical to keep acquiring new knowledge in today’s ever changing landscape of both world order and opportunities available to professionals.” - Joanna Zarach

“The quality and pace of the videos and material is top-notch. I really like having different instructors for different modules and having two instructors interacting together makes the material more vivid and entertaining.” - Miguel Hurtado

“Armed with the knowledge I have gained from his course, I can introduce my team to certain methods that can be applied to our day to day work.” - Anonymous Learner
Module 1: Making sense of unstructured data

Case Study 1: Genetic Codes
- **Case Study Activity Description:** Use K-means to figure out that DNA is composed of three-letter words. We'll help by demonstrating how to apply data visualization to genomic sequence analysis.
- **Data Sets & format:** DNA text string

Case Study 2: LDA Analysis
- **Case Study Activity Description:** Find themes in project descriptions using LDA. We'll help by giving you tips on how to do your own analysis on MIT EECS faculty data using stochastic variational inference on LDA.
- **Data Sets & format:** Scrape your own

Case Study 3: PCA: Identifying Faces
- **Case Study Activity Description:** Implement your own image classification algorithm that helps classify photos of people's faces. We'll help by giving you tips on how to use PCA, along with examples and pseudo-code for the programming environment.
- **Data Sets & format:** Instructors photos provided (14). Any other images will work, as long as they obey the restrictions noted in the Self Help document.

Case Study 4: Spectral Clustering: Grouping News Stories
- **Case Study Activity Description:** Build your own clustering for online news stories—similar to how Google News organizes stories via auto-generated topics. We'll help by giving you tips on Spectral Clustering, along with examples and pseudo-code for the programming environment.
- **Data Sets & format:** Instructions for downloading news stories off the web.
Module 2: Making sense of unstructured data

Case Study 1: Predicting Wages 1
- **Case Study Activity Description:** Predict wages and assess predictive performance using various characteristics of workers. We’ll help by describing the wage prediction model.
- **Data Sets & format:** CPS 2012 Data, Rdata format

Case Study 2: Gender Wage Gap
- **Case Study Activity Description:** Estimate the difference in predicted wages between men and women with the same job characteristics. We’ll help by describing the estimation technique and presenting the results.
- **Data Sets & format:** CPS 2012 Data, Rdata format

Case Study 3: Do Poor Countries Grow Faster than Rich Countries?
- **Case Study Activity Description:** Use a large dimensional dataset to answer the question: Do poor countries grow faster than rich countries? We’ll help by describing the estimation technique, giving you the tools, and presenting the results.
- **Data Sets & format:** Barro-Lee Growth Data. Rdata format.

Case Study 4: Predicting Wages 2
- **Case Study Activity Description:** Predict wages using several machine learning methods and splitting data. We’ll help by describing the estimation technique and presenting the results.
- **Data Sets & format:** 2015 CPS data, Rdata format.

Case Study 5: The Effect of Gun Ownership on Homicide Rate
- **Case Study Activity Description:** Use machine learning methods to estimate the effect of gun ownership on the homicide rate. We’ll help by describing the estimation technique and presenting the results.
- **Data Sets & format:** U.S. Census Bureau Dataset. Csv format.
MODULE 3.1: Classification and Hypothesis Testing

Case-study 1: Logistic Regression: The Challenger Disaster

- **Case Study Activity Description:** Learn how to apply Logistic Regression in a practical real-world setting. We’ll help by giving you tips, examples, and pseudo-code for the programming environments.
- **Data Sets & format:** Made available as a csv file along with the case study.

MODULE 3.2: Deep Learning

Case Study 2: Decision boundary of a deep neural network

- **Case Study Activity Description:** Play with one or two layer perceptrons to assess their decision boundaries. We’ll help by explaining the multiple dimensions of perceptrons.
- **Data Sets & format:** Synthetic 2D data points.

Case Study 1: Recommending Movies

- **Case Study Activity Description:** Build your own recommendation system for movies like the one used by Netflix. We’ll help by giving you tips, examples, and pseudo-code for the programming environments.
- **Data Sets & format:** MovieLens dataset - public set

Case Study 2: Recommend New Songs to Users Based on Their Listening Habits

- **Case Study Activity Description:** Build your own recommendation system for songs like the one used by Spotify. We’ll help by giving you tips, examples, and pseudo-code for the programming environments.
- **Data Sets & format:** Million Song dataset
Case Study 3: Make New Product Recommendations

- **Case Study Activity Description:** Build your own recommendation system for products on an e-commerce website like the one used by Amazon.com. We'll help by giving you tips, examples, and pseudo-code for the programming environments.
- **Data Sets & format:** Amazon Reviews data

MODULE 5: Networks and Graphical Models

Case study 1: Navigation / GPS

1.1: Kalman Filtering: Tracking the 2D Position of an Object when moving with Constant Velocity

- **Case Study Activity Description:** Generate data, build the model for the motion dynamics, and perform the Kalman Filtering algorithm. We'll help by giving you tips, examples, and pseudo-code for the programming environment.
- **Data Sets & format:** Generating your own data. Model explanation and other parameter details provided in a separate write-up.

1.2: Kalman Filtering: Tracking the 3D Position of an Object falling due to gravity.

- **Case Study Activity Description:** Generate data, build the model for the motion dynamics, perform the Kalman Filtering algorithm. We'll help by giving you tips, examples, and pseudo-code for the programming environment.
- **Data Sets & format:** Generating your own data. Model explanation and other parameter details provided in a separate write-up.
**Case Study 2: Identifying New Genes that cause Autism**
- **Case Study Activity Description:** Use network-theoretic ideas to identify new candidate genes that might cause autism. We’ll help by giving you tips, examples, and pseudo-code for the programming environment.
- **Data Sets & format:** Made available as csv files.

**MODULE 6: Case studies**

**Case study 1: New York City**
- **Case Study Description:** To predict the trip duration of a new york taxi cab ride, build different types of features and evaluate them. We will start by describing what a feature is in this context, then develop some very simple features and add features using the software package featuretools. We will assess how these features perform in predicting trip duration.
- **Datasets and format:** Multiple csv files, loaded as pandas data frames.

**Case study 2: Prediction Engineering Using UK Retail Dataset**
- **Case Study Description:** Given a retail dataset we will formulate a prediction problem as a retailer would, and develop an end-to-end solution using featuretools for feature engineering and scikit learn for modeling. We will change the prediction problem and tunes its parameters and see how the model performance changes.
- **Datasets and format:** Multiple csv files, loaded as pandas data frames.
INSTRUCTORS

In this course, you won't just discover new strategies, tools, and insights— you'll put them to the test. Every course module features a selection of case studies and hands-on projects that help you apply your newfound knowledge to realistic business challenges.

Devavrat Shah  
Course Co-Director

Philippe Rigollet  
Course Co-Director

Guy Bresler  
Course Co-Director

Ankur Moitra  
Course Co-Director

Caroline Uhler  
Course Co-Director

Victor Chernozhukov  
Course Co-Director

David Gamarnik  
Course Co-Director

Stefanie Jegelka  
Course Co-Director

Jonathan Kelner  
Course Co-Director

Ankur Moitra  
Course Co-Director

Kaylan Veeramachaneni  
Course Co-Director
Contact Us

ppp.edge@pearson.com
https://in.pearson.com/ppp

0124 - 4185100

Pearson Professional Programs,
Vatika First Indian Place, Tower C,
Ground Floor (Near M.G. Road
Metro Station), Gurugram, India