MASSACHUSETTS INSTITUTE OF TECHNOLOGY

IDSS
MIT INSTITUTE FOR DATA, SYSTEMS, AND SOCIETY

DATA SCIENCE AND MACHINE LEARNING: MAKING DATA-DRIVEN DECISIONS

Become a data-driven decision maker with the 12-week online program delivered by MIT faculty

ABOUT

MIT IDSS

Education and research at MIT Institute for Data, Systems, and Society (IDSS) are undertaken with the aim to provide solutions to complex societal challenges by understanding and analyzing data. The institute is thus committed to the development of analytical methods that can be applied to diverse areas such as finance, energy systems, urbanization, social networks, and health.

MIT IDSS embraces the collision and synthesis of ideas and methods from analytical disciplines including statistics, data science, information theory and inference, systems and control theory, optimization, economics, human and social behavior, and network science. These disciplines are relevant both for understanding complex systems and for presenting design principles and architectures that allow for the systems' quantification and management.

MISSION

The mission of MIT IDSS is to advance education and research in state-of-the-art analytical methods in information and decision systems, statistics and data science, and the social sciences, and to apply these methods to address complex societal challenges in a diverse set of areas such as finance, energy systems, urbanization, social networks, and health.

Technology advances in areas such as smart sensors, big data, communications, computing, and social networking are rapidly scaling the size and complexity of interconnected systems and networks, and at the same time are generating masses of data that can lead to new insights and understanding. Research at IDSS aims to understand and analyze data from across these systems, which present unique and substantial challenges due to scale, complexity, and the difficulties of extracting clear, actionable insights.

ABOUT

THE PROGRAM

Demand for professionals skilled in data, analytics, and machine learning is exploding. According to a report by the U.S. Bureau of Labor Statistics, the demand for data science is set to increase, creating 11.5 million new data-driven jobs by 2026. Data scientists bring value to organizations across industries because they are able to solve complex challenges with data and drive important decision-making processes.

The MIT Institute for Data, Systems, and Society (IDSS) understands the power of uncovering the true value of your data and has created a variety of online courses and programs to take your data analytics skills to the next level. Whether you are looking to break into the field, seeking career development opportunities, or simply want to provide more valuable insights to your company, there offerings will teach you to harness data in new and innovative ways.

PROGRAM

BENEFITS

- Learn online from 11 award-winning MIT faculty and instructors
- Fuel your career transition with CV & Linkedin review sessions, and demonstrate your data science expertise by building a portfolio of 3 real-life projects and 15+ case studies
- Work in a robust collaborative environment to network with peers in Data Science and Machine Learning
- Get a Certificate of Completion by MIT IDSS
- Get mentorship from industry experts on the applications of concepts taught by the faculty

PROGRAM

STRUCTURE

30+ hours of recorded lectures from MIT faculty and mentored sessions with industry experts, along with hands-on projects and conceptual sessions, structured as follows:

WEEK 1 & 2	**Foundational courses:** Python and Statistics for Data Science
WEEK 3	**Clustering, Spectral Clustering, Components, and Embeddings:** Sessions, hands-on case studies, and quizzes
WEEK 4	**Learning Break** **Conceptual Session:** Making sense of Unstructured Data and Regression and Prediction
WEEK 5 & 6	**Supervised and Unsupervised Learning:** Regression, Hypothesis Testing, and Classification
WEEK 7	**Learning Break** **Conceptual Session:** Classification and Hypothesis Testing
WEEK 8 & 9	**Deep Learning and Recommendation Systems:** Neural Networks, Collaborative Filtering, and personalized recommendation techniques through sessions, hands-on case studies, and quizzes
WEEK 10	**Learning Break** **Conceptual Session:** Deep Learning and Recommendation Systems
WEEK 11 & 12	**Networks and Predictive Analytics:** Graphical models, Predictive Modeling, and feature engineering techniques through sessions, hands-on case studies, and quizzes

Note: Conceptual sessions help learners solidify concepts learned in the video content of the week or as a preparatory session for the upcoming courses. This also helps learners experience extended QnA with an industry expert in the webinar.

WHO IS THIS
PROGRAM FOR?

- Data Scientists, Data Analysts, and working professionals who wish to extract actionable insights from large volumes of data
- Due to the broad nature of the program, it is suited for both early career professionals and senior managers, including technical managers, business intelligence analysts, data science managers, data science enthusiasts, IT practitioners, management consultants, and business managers
- Those with some academic/professional training in applied mathematics/statistics will find it easier to learn. However, all participants can make the most out of this program by learning these topics through Great Learning's ample support

AFTER THIS COURSE,
YOU WILL BE ABLE TO

- Understand the intricacies of Data Science techniques and their applications to real-world problems
- Implement various Machine Learning techniques to solve complex problems and make data-driven business decisions
- Explore two major realms of Machine Learning, Deep Learning, and Neural Networks, and how they can be applied to areas such as Computer Vision
- Choose how to represent your data when making predictions
- Understand the theory behind recommendation systems and explore their applications to multiple industries and business contexts
- Build an industry-ready portfolio of projects to demonstrate your ability to extract business insights from data

PROGRAM

CURRICULUM

The program is 12 weeks long:

WEEK 1-2

Foundational courses: Python and Statistics for Data Science

Python for Data Science

- Numpy
- Pandas
- Data Visualization

Case Study 1: *FIFA World Cup analysis*
Assessment: *Movielens project*

Stats for Data Science

- Descriptive Statistics
- Inferential Statistics

Case Study 2: *Fitness product customer footfall analysis*
Assessment: *Movielens project*

WEEK 3

Making sense of Unstructured Data

Introduction

- What is unsupervised learning, and why is it challenging?
- Examples of unsupervised learning

Clustering

- What is Clustering?
- When to use Clustering
- K-means Preliminaries
- The K-means algorithm
- How to evaluate Clustering
- Beyond K-means: What really makes a Cluster?
- Beyond K-means: Other notions of distance
- Beyond K-means: Data and pre-processing
- Beyond K-means: Big data and Nonparametric Bayes
- Beyond Clustering

Case Study 1: *Genetic Codes*
Case Study 2: *Finding themes in the project description*

Spectral Clustering, Components, and Embeddings

- What if we do not have features to describe the data or not all are meaningful?
- Finding the principal components in data and applications
- The magic of Eigenvectors I
- Clustering in Graphs and Networks
- Features from graphs: The magic of Eigenvectors II
- Spectral Clustering
- Modularity Clustering
- Embeddings: New features and their meaning

Case Study 3: *PCA: Identifying faces*
Case Study 4: *Spectral Clustering: Grouping news stories*

WEEK 4

LEARNING BREAK

With a conceptual session on making sense of Unstructured Data

WEEK 5

Regression and Prediction

Classical Linear and Nonlinear Regression and Extensions

- Linear Regression with one and several variables
- Linear Regression for prediction
- Linear Regression for causal inference
- Logistic and other types of Nonlinear Regression

Case Study 1: *Predicting Wages 1*
Case Study 2: *Gender Wage Gap*

Modern Regression with High-Dimensional Data

- Making good predictions with high-dimensional data; avoiding
- Overfitting by validation and cross-validation
- Regularization by Lasso, Ridge, and their modifications
- Regression Trees, Random Forest, Boosted Trees

Case Study 3: *Do poor countries grow faster than rich countries?*

The Use of Modern Regression for Causal Inference

- Randomized Control Trials
- Observational Studies with Confounding

Case Study 4: *Predicting wages 2*
Case Study 5: *The effect of gun ownership on homicide rates*

WEEK 6

Classification and Hypothesis Testing

- What are anomalies? What is fraud? Spams?
- Binary Classification: False Positive/Negative, Precision/Recall, F1-Score
- Logistic and Probit Regression: Statistical Binary Classification
- Hypothesis Testing: Ratio Test and Neyman-Pearson p-values: Confidence
- Support Vector Machine: Non-statistical Classifier
- Perceptron: Simple Classifier with elegant interpretation

Case Study 1: *Logistic Regression: The Challenger Disaster*

WEEK 7

LEARNING BREAK

With a conceptual session on Classification and Hypothesis Testing

WEEK 8

Deep Learning

- What is Image Classification? Introduce ImageNet and show examples
- Classification using a single linear threshold (perceptron)
- Hierarchical representations
- Fitting parameters using back-propagation
- Non-convex functions
- How interpretable are its features?
- Manipulating Deep Nets (Ostrich Example)
- Transfer Learning
- Other applications I: Speech Recognition
- Other applications II: Natural Language Processing

Case Study 1: *Decision Boundary of a Deep Neural Network*

WEEK 9

Recommendation Systems

Recommendations and Ranking

- What does a Recommendation System do?
- What is the Recommendation Prediction Problem? And what data do we have?
- Using Population Averages
- Using Population Comparisons and Ranking

Case Study 1: *Recommending movies*

Collaborative Filtering

- Personalization using collaborative filtering using similar users
- Personalization using collaborative filtering using similar items
- Personalization using collaborative filtering using similar users and items

Case Study 2: *Recommend new songs to the users based on their listening habits*

Personalized Recommendations

- Personalization using Comparisons, Rankings, and Users-items
- Hidden Markov Model / Neural Nets, Bipartite graph, and Graphical Model
- Using side-information
- 20 questions and active learning
- Building a system: Algorithmic and system challenges

Case Study 3: *Make new product recommendations*

WEEK 10

LEARNING BREAK

With a conceptual session on Deep Learning and Recommendation Systems

WEEK 11

Networking and Graphical Models

Introduction

- Introduction to networks
- Examples of networks
- Representation of networks

Case Study 1: *Navigation / GPS*
1.1: *Kalman Filtering: Tracking the 2D position of an object when moving with constant velocity*
1.2: *Kalman Filtering: Tracking the 3D position of an object falling due to gravity.*

Networks

- Centrality measures: degree, eigenvector, and page-rank
- Closeness and betweenness centrality
- Degree distribution, clustering, and small world
- Network Models: Erdos-Renyi, configuration model, preferential attachment
- Stochastic Models on networks for the spread of viruses or ideas
- Influence maximization

Case Study 2: *Identifying new genes that cause autism*

Graphical Models

- Undirected Graphical Models
- Ising and Gaussian Models
- Learning Graphical Models from data
- Directed graphical models
- V-structures, “explaining away,” and learning Directed Graphical Models
- Inference in Graphical Models: Marginals and message passing
- Hidden Markov Model (HMM)
- Kalman Filter

WEEK 12

Predictive Analytics

Predictive Modeling for Temporal Data

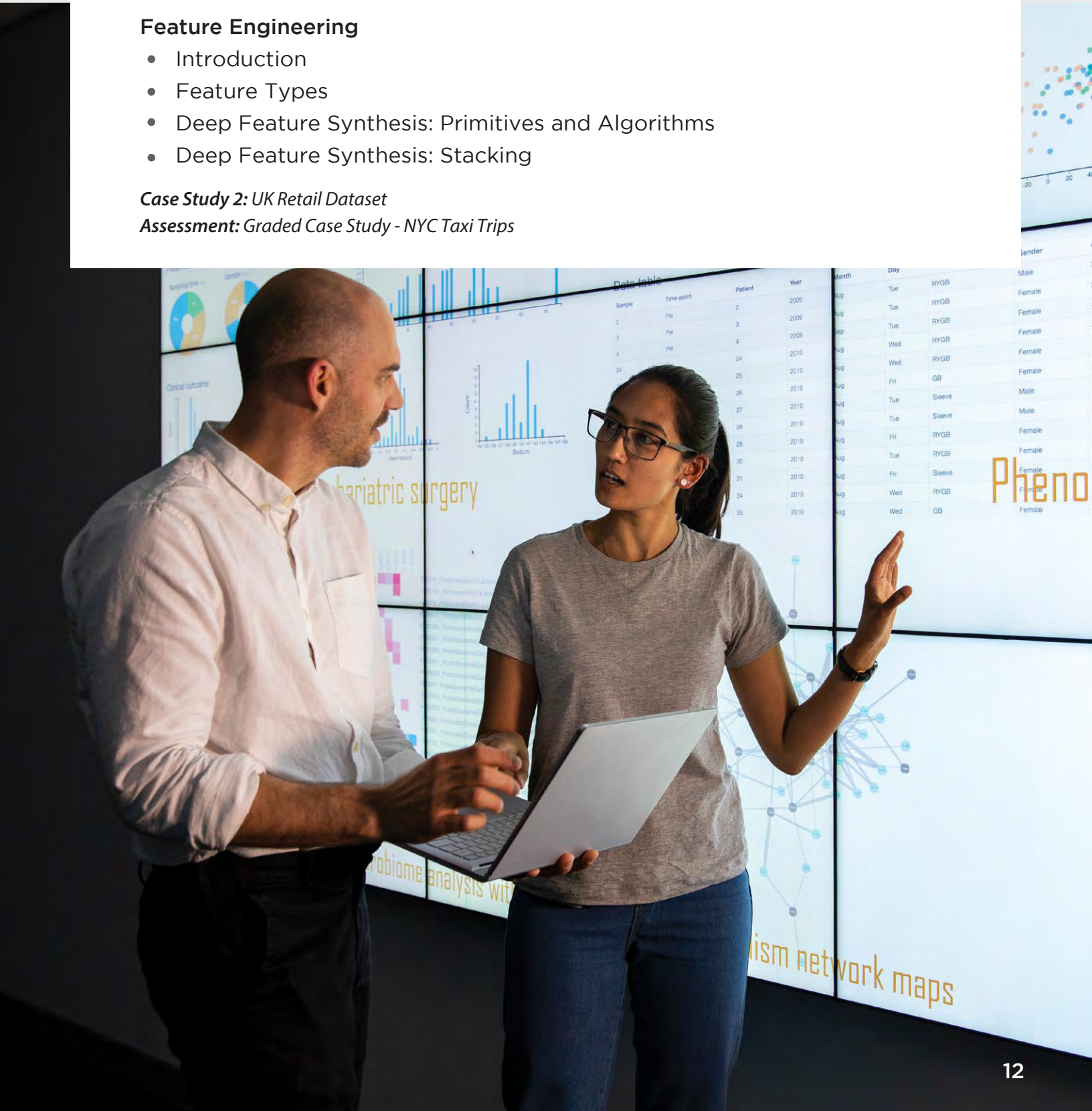
- Prediction Engineering

Case Study 1: *NYC Taxi*

Feature Engineering

- Introduction
- Feature Types
- Deep Feature Synthesis: Primitives and Algorithms
- Deep Feature Synthesis: Stacking

Case Study 2: *UK Retail Dataset*
Assessment: *Graded Case Study - NYC Taxi Trips*



FACULTY

Devavrat Shah
Professor, EECS and IDSS, MIT

Philippe Rigollet
Professor, Mathematics and IDSS, MIT

Caroline Uhler
Henry L. & Grace Doherty Associate Professor,
EECS and IDSS, MIT

Victor Chernozhukov
Professor, Economics and IDSS, MIT

Stefanie Jegelka
X-Consortium Career Development Associate Professor,
EECS and IDSS, MIT

Ankur Moitra
Rockwell International Career Development Associate Professor, Mathematics and IDSS, MIT

Tamara Broderick
Associate Professor, EECS and IDSS, MIT

David D. Gamarnik
Nanyang Technological University Professor of Operations Research, Sloan School of Management and IDSS, MIT

Jonathan Kelner
Professor, Mathematics, MIT

Kalyan Veeramachaneni
Principal Research Scientist at the Laboratory for Information and Decision Systems, MIT

Guy Bresler
Associate Professor, EECS and IDSS, MIT

PROGRAM

MENTORS

The program coaches you to work on hands-on industry-relevant projects by Data Science and Machine Learning experts via live and personalized mentored learning sessions to give you a practical understanding of core concepts. A few of the industry experts engaged with us as program mentors include:

Matthew Nickens
Manager, Partnership Science
Meta

Udit Mehrotra
Data Scientist
Dell

Bhaskarjit Sarmah
Data Scientist
BlackRock

Shirish Gupta
Lead Data Scientist
Novartis

Bradford Tuckfield
Senior Manager, Data Science
Charles Schwab

Vaibhav Verdhan
Analytics Leader,
Global Advanced Analytics
AstraZeneca

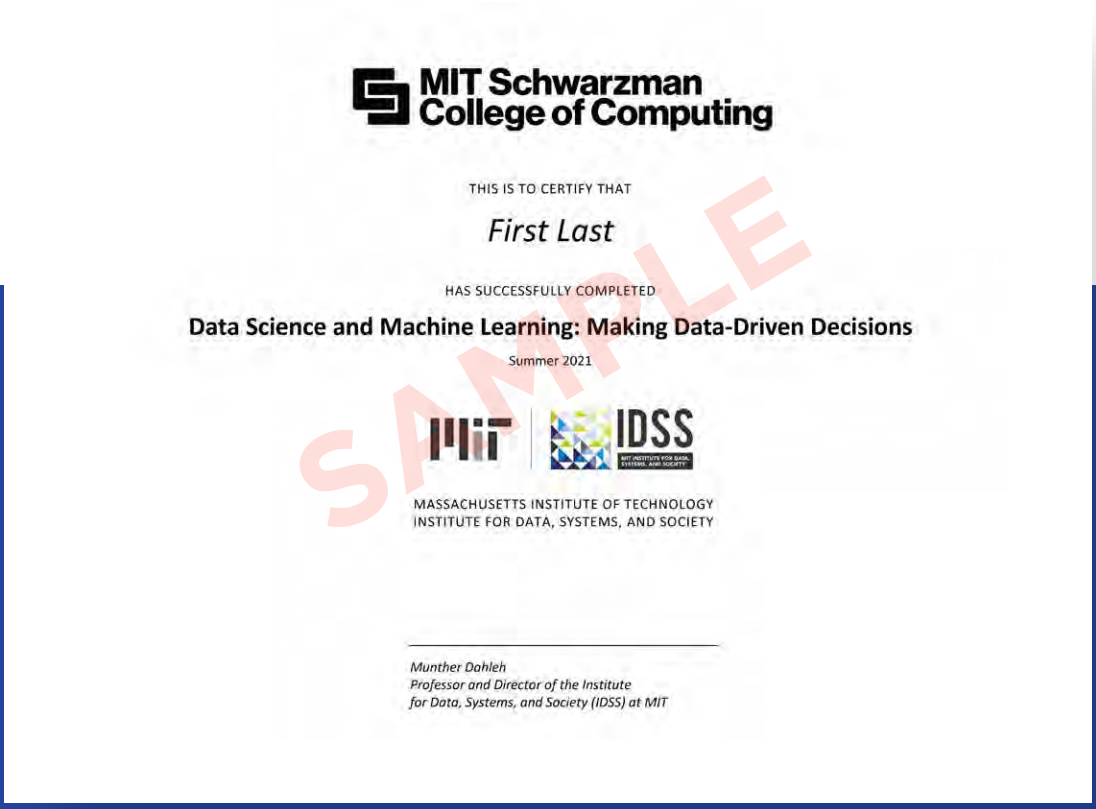
Odaiyappan Pichappan
Senior Data Scientist
Verizon 4G Wireless

PROGRAM MANAGER:

YOUR PERSONAL GUIDE

Your dedicated Program Manager, provided by Great Learning, will be your single point of contact for all academic and non-academic queries in the program. They will keep track of your learning journey, give you personalized feedback, and the required nudges to ensure your success.

CERTIFICATE



The image is for illustrative purposes only. The actual certificate may be subject to change at the discretion of the university.

APPLICATION PROCESS

Step-1

Application Form

Register by completing the online application form.

Step-2

Application Screening

Your application will be reviewed to determine whether you're eligible for this program.

Step-3

Join the Program

If selected, you will receive an offer for the upcoming cohort. Secure your seat by paying the fee.

APPLICATION & FEE DETAILS

Program Duration: **12 weeks**

Fees: **USD 1,900**

Start Date: **May 14, 2022**

MIT IDSS Data Science and Machine Learning Program, with curriculum developed and taught by MIT faculty, is delivered in collaboration with

Great Learning
POWER AHEAD

Great Learning is an ed-tech platform with a mission to make professionals proficient and future-ready. Its programs always focus on the next frontier of growth in the industry and currently straddle across Analytics, Data Science, Big Data, Machine Learning, Artificial Intelligence, Deep Learning, Cloud Computing, and more. Great Learning uses technology, high-quality content, and industry collaborations to deliver an immersive learning experience that helps candidates learn, apply, and demonstrate their competencies. All programs are offered in collaboration with leading global universities and are taken by thousands of professionals every year from 160+ countries.

READY TO BECOME A DATA-DRIVEN DECISION MAKER?

APPLY NOW

SPEAK TO A PROGRAM ADVISOR

+1 617 539 7216

HAVE QUESTIONS ABOUT THE PROGRAM OR HOW IT FITS IN WITH YOUR CAREER GOALS?

dsml.mit@mygreatlearning.com

VISIT OUR WEBSITE

https://www.mygreatlearning.com/mit-data-science-machine-learning-program/