



Post Graduate Program in Data Science

In collaboration with IBM

simplilearn

Table of Contents

About the Program	03
The Key Features of the Post Graduate Program in Data Science in partnership with Purdue University	04
About the Post Graduate Program in Data Science in Partnership with Purdue University	05
About Simplilearn	05
Program Eligibility Criteria and Application Process	06
Learning Path Visualization	08
Program Outcomes	09
Who Should Enroll in this Program	10
Courses	
● Step 1 - Programming Refresher	11
● Step 2 - Statistics Essential for Data Science	12
● Step 3 - R Programming for Data Science	14
● Step 4 - Data Science with R	15
● Step 5 - Python for Data Science	17
● Step 6 - Data Science with Python	18
● Step 7 - Machine Learning	20
● Step 8 - Tableau	22
● Step 9 - Natural Language Processing	24
● Step 10 - Data Science Capstone Project	25
Electives	26
Certification	27
Advisory Board Members	28



About the Program

Accelerate your career with this acclaimed Post Graduate Program in Data Science, in partnership with Purdue University and in collaboration with IBM, and featuring the perfect mix of theory, case studies and extensive hands-on practicum. In partnership with Purdue University, this program is a comprehensive Data Science education – leveraging Purdue's academic excellence in Data Science and Simplilearn's partnership with IBM.

Designed to give recent graduates and experienced professionals an extensive Data Science education, this Post Graduate Program is a blend of online self-paced videos, live virtual classes, hands-on projects, and labs, with mentorship sessions to provide a high-engagement learning experience and real-world applications to help you master essential Data Science skills. This program offers in-depth exposure to technologies including R, Python, Machine Learning, Tableau, Natural Language Processing, and prepares you for an exciting career in Data Science.

The Key Features of the Post Graduate Program in Data Science in partnership with Purdue University



Purdue Post Graduate Program certification



Purdue Alumni Association Membership



International recognition by Purdue University



Leading industry-recognized analytics course



Program in collaboration with IBM



Industry-recognized IBM certificates



25+ hands-on projects and one capstone



Limited class size for an optimal experience



300+ hours of Blended Learning

About the Post Graduate Program in Data Science in partnership with Purdue University

Purdue University, a top public research institution, offers higher education at its highest proven value. Committed to affordability, the University has frozen tuition and most fees at 2012-13 levels. Committed to student success, Purdue is changing the student experience with a greater focus on faculty-student interaction and creative use of technology. Committed to pursuing scientific discoveries and engineered solutions, Purdue has streamlined pathways for faculty and student innovators who have a vision for moving the world forward.

This Data Science Post Graduate Program in partnership with Purdue University will open pathways for your career in virtually every realm of business—from healthcare to education to manufacturing.

Upon successfully completing this program, you will:

- › Receive a joint Purdue-Simplilearn certificate of completion
- › Be eligible able to join the Purdue alumni association and participate in its various networking opportunities and career events



About Simplilearn

Simplilearn is a leader in digital skills training, focused on the emerging technologies that are transforming our world. Our unique blended learning approach drives learner engagement and backed by the industry's highest course completion rates. Partnering with professionals and companies, we identify their unique needs and provide outcome-centric solutions to help them achieve their professional goals.

Program Eligibility Criteria and Application Process

Those wishing to enroll in the Post Graduate Program in Data Science in partnership with Purdue University will be required to apply for admission to the program.

Eligibility Criteria

For admission to this Post Graduate Program in Data Science, candidates should have:

- ✓ A bachelor's degree with an average of 50% or higher marks
- ✓ Basic understanding of programming concepts and mathematics
- ✓ Working Professionals with 2+ years of experience are preferred to apply for this program

Application Process

The application process consists of three simple steps. An offer of admission will be made to the selected candidates and accepted by the candidates by paying the admission fee.

STEP 1

SUBMIT AN APPLICATION

Complete the application and include a brief statement of purpose telling our admissions counselors why you're interested and qualified to be part of the Post Graduate Program in Data Science.

STEP 2

APPLICATION REVIEW

After you submit your application, a panel of admissions counselors will review your application and statement of purpose to determine your qualifications and interest in the program.

STEP 3

ADMISSION

An offer of admission will be made to qualified candidates, and you can accept this offer by paying the program fee.

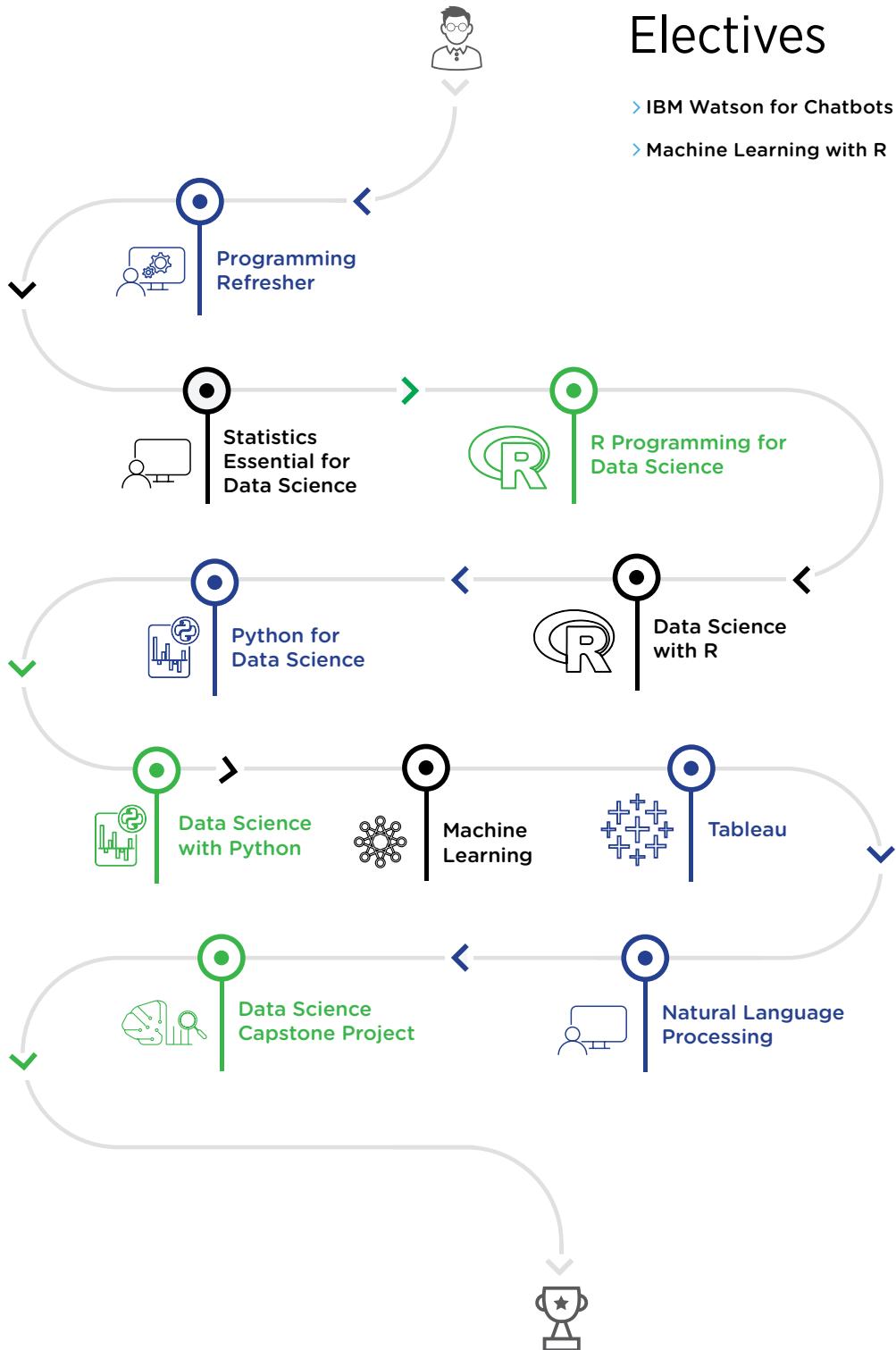


Talk to an Admissions Counselor

We have a team of dedicated admissions counselors who are here to help guide you in applying to the program. They are available to:

- ✓ Address questions related to the application
- ✓ Assist with financial aid (if required)
- ✓ Help you resolve your questions and understand the program

Learning Path



Program Outcomes



Gain an in-depth understanding of data structure and data manipulation



Understand and use linear and non-linear regression models and classification techniques for data analysis



Obtain a comprehensive knowledge of supervised and unsupervised learning models such as linear regression, logistic regression, clustering, dimensionality reduction, K-NN and pipeline



Perform scientific and technical computing using the SciPy package and its sub-packages such as Integrate, Optimize, Statistics, IO, and Weave



Gain expertise in mathematical computing using the NumPy and Scikit-Learn package



Master the concepts recommendation engine, time series modeling, gain practical mastery over principles, algorithms, and applications of Machine Learning



Learn to analyze data using Tableau and become proficient in building interactive dashboards



Understand deep reinforcement learning techniques applied in Natural Language Processing



Understand the different components of the Hadoop ecosystem and learn to work with HBase, its architecture and data storage, learning the difference between HBase and RDBMS, and use Hive and Impala for partitioning



Understand MapReduce and its characteristics, plus learn how to ingest data using Sqoop and Flume



Who Should Enroll in this Program?

This program caters to working professionals from a variety of industries and backgrounds; the diversity of our students adds richness to class discussions and interactions.

The Data Science role requires an amalgam of experience, Data Science knowledge, and using the correct tools and technologies. It is a solid career choice for both new and experienced professionals. Aspiring professionals of any educational background with an analytical

frame of mind are most suited to pursue this Post Graduate Program in Data Science, including:

- ✓ IT professionals
- ✓ Analytics managers
- ✓ Business analysts
- ✓ Software developers
- ✓ Beginners or recent graduates with bachelor's or master's degree

Programming Refresher

Programming is an increasingly important skill; this course will establish your proficiency in handling basic programming concepts. By the end of this program, you will understand object-oriented programming; basic programming concepts such as data types, variables, strings, loops, and functions; and software engineering concepts such as multithreading and multitasking using Python.

Key Learning Objectives

- ✓ Obtain fundamental knowledge of programming basics
- ✓ Achieve an understanding of object-oriented programming principles including data types, variables, strings, loops, and functions
- ✓ Comprehend software engineering concepts, such as multithreading and multitasking using Python

Course curriculum

- ✓ Lesson 01 - Course Introduction
- ✓ Lesson 02 - Programming Basics

Statistics Essential for Data Science

Statistics is the science of assigning a probability through the collection, classification, and analysis of data. A foundational part of Data Science, this course will enable you to define statistics and essential terms related to it, explain measures of central tendency and dispersion, and comprehend skewness, correlation, regression, distribution. You will be able to make data-driven predictions through statistical inference.

Key Learning Objectives

- ✓ Understand the fundamentals of statistics
- ✓ Work with different types of data
- ✓ How to plot different types of data
- ✓ Calculate the measures of central tendency, asymmetry, and variability
- ✓ Calculate correlation and covariance
- ✓ Distinguish and work with different types of distribution
- ✓ Estimate confidence intervals
- ✓ Perform hypothesis testing
- ✓ Make data-driven decisions
- ✓ Understand the mechanics of regression analysis
- ✓ Carry out regression analysis
- ✓ Use and understand dummy variables
- ✓ Understand the concepts needed for Data Science even with Python and R

Course curriculum

- ✓ Lesson 01 - Introduction
- ✓ Lesson 02 - Sample or Population Data?
- ✓ Lesson 03 - The Fundamentals of Descriptive Statistics
- ✓ Lesson 04 - Measures of Central Tendency, Asymmetry, and Variability
- ✓ Lesson 05 - Practical Example: Descriptive Statistics
- ✓ Lesson 06 - Distributions
- ✓ Lesson 07 - Estimators and Estimates
- ✓ Lesson 08 - Confidence Intervals: Advanced Topics
- ✓ Lesson 09 - Practical Example: Inferential Statistics
- ✓ Lesson 10 - Hypothesis Testing: Introduction
- ✓ Lesson 11 - Hypothesis Testing: Let's Start Testing!
- ✓ Lesson 12 - Practical Example: Hypothesis Testing
- ✓ Lesson 13 - The Fundamentals of Regression Analysis
- ✓ Lesson 14 - Subtleties of Regression Analysis
- ✓ Lesson 15 - Assumptions for Linear Regression Analysis
- ✓ Lesson 16 - Dealing with Categorical Data
- ✓ Lesson 17 - Practical Example: Regression Analysis

R Programming for Data Science

Gain insight into the R Programming language with this introductory course. An essential programming language for data analysis, R Programming is a fundamental key to becoming a successful Data Science professional. In this course, you will learn how to write R code, learn about R's data structures, and create your own functions. After the completion of this course, you will be fully able to begin your first data analysis.

Key Learning Objectives

- ✓ Learn about math, variables, and strings, vectors, factors, and vector operations
- ✓ Gain fundamental knowledge on arrays and matrices, lists, and data frames
- ✓ Get understanding on conditions and loops, functions in R, objects, classes, and debugging
- ✓ Learn how to accurately read text, CSV, and Excel files, plus how to write and save data objects in R to a file
- ✓ Understand and work on strings and dates in R

Course curriculum

- ✓ Lesson 01 - R Basics
- ✓ Lesson 02 - Data Structures in R
- ✓ Lesson 03 - R Programming Fundamentals
- ✓ Lesson 04 - Working with Data in R
- ✓ Lesson 05 - Strings and Dates in R

Data Science with R

The next step to becoming a data scientist is learning R—the most in-demand open source technology. R is a powerful Data Science and analytics language, which has a steep learning curve and a very vibrant community. This is why it is quickly becoming the technology of choice for organizations who are adopting the power of analytics for competitive advantage.

Key Learning Objectives

- ✓ Gain a foundational understanding of business analytics
- ✓ Install R, R-studio, and workspace setup, and learn about the various R packages
- ✓ Master R programming and understand how various statements are executed in R
- ✓ Gain an in-depth understanding of data structure used in R and learn to import/export data in R
- ✓ Define, understand and use the various apply functions and DPLYR functions
- ✓ Understand and use the various graphics in R for data visualization
- ✓ Gain a basic understanding of various statistical concepts
- ✓ Understand and use hypothesis testing method to drive business decisions
- ✓ Understand and use linear, non-linear regression models, and classification techniques for data analysis
- ✓ Learn and use the various association rules and Apriori algorithm
- ✓ Learn and use clustering methods including K-means, DBSCAN, and hierarchical clustering

Course curriculum

- ✓ Lesson 01 - Introduction to Business Analytics
- ✓ Lesson 02 - Introduction to R Programming
- ✓ Lesson 03 - Data Structures
- ✓ Lesson 04 - Data Visualization
- ✓ Lesson 05 - Statistics for Data Science I
- ✓ Lesson 06 - Statistics for Data Science II
- ✓ Lesson 07 - Regression Analysis
- ✓ Lesson 08 - Classification
- ✓ Lesson 09 - Clustering
- ✓ Lesson 10 - Association

Python for Data Science

Kickstart your learning of Python for Data Science with this introductory course and familiarize yourself with programming. Carefully crafted by IBM, upon completion of this course you will be able to write your Python scripts, perform fundamental hands-on data analysis using the Jupyter-based lab environment, and create your own Data Science projects using IBM Watson.

Key Learning Objectives

- ✓ Write your first Python program by implementing concepts of variables, strings, functions, loops, conditions
- ✓ Understand the nuances of lists, sets, dictionaries, conditions and branching, objects and classes
- ✓ Work with data in Python such as reading and writing files, loading, working, and saving data with Pandas

Course curriculum

- ✓ Lesson 01 - Python Basics
- ✓ Lesson 02 - Python Data Structures
- ✓ Lesson 03 - Python Programming Fundamentals
- ✓ Lesson 04 - Working with Data in Python
- ✓ Lesson 05 - Working with NumPy Arrays

Data Science with Python

This Data Science with Python course will establish your mastery of Data Science and analytics techniques using Python. In this Python for Data Science course, you'll learn the essential concepts of Python programming and gain in-depth knowledge in data analytics, Machine Learning, data visualization, web scraping, and natural language processing. Python is a required skill for many Data Science positions, so jump-start your career with this interactive, hands-on course.

Key Learning Objectives

- ✓ Gain an in-depth understanding of Data Science processes, data wrangling, data exploration, data visualization, hypothesis building, and testing
- ✓ Install the required Python environment and other auxiliary tools and libraries
- ✓ Understand the essential concepts of Python programming such as data types, tuples, lists, dicts, basic operators and functions
- ✓ Perform high-level mathematical computing using the NumPy package and its vast library of mathematical functions
- ✓ Perform scientific and technical computing using the SciPy package and its sub-packages such as Integrate, Optimize, Statistics, IO, and Weave
- ✓ Perform data analysis and manipulation using data structures and tools provided in the Pandas package
- ✓ Gain expertise in Machine Learning using the Scikit-Learn package
- ✓ Gain an in-depth understanding of supervised learning and unsupervised learning models such as linear regression, logistic regression, clustering, dimensionality reduction, K-NN and pipeline

- ✓ Use the Scikit-Learn package for natural language processing
- ✓ Use the matplotlib library of Python for data visualization
- ✓ Extract useful data from websites by performing web scraping using Python
- ✓ Integrate Python with Hadoop, Spark, and MapReduce

Course curriculum

- ✓ Lesson 01 - Data Science Overview
- ✓ Lesson 02 - Data Analytics Overview
- ✓ Lesson 03 - Statistical Analysis and Business Applications
- ✓ Lesson 04 - Python Environment Setup and Essentials
- ✓ Lesson 05 - Mathematical Computing with Python (NumPy)
- ✓ Lesson 06 - Scientific Computing with Python (Scipy)
- ✓ Lesson 07 - Data Manipulation with Pandas
- ✓ Lesson 08 - Machine Learning with Scikit-Learn
- ✓ Lesson 09 - Natural Language Processing with Scikit Learn
- ✓ Lesson 10 - Data Visualization in Python using Matplotlib
- ✓ Lesson 11 - Web Scraping with BeautifulSoup
- ✓ Lesson 12 - Python Integration with Hadoop MapReduce and Spark

Machine Learning

Simplilearn's Machine Learning course will make you an expert in Machine Learning, a form of Artificial Intelligence that automates data analysis to enable computers to learn and adapt through experience to do specific tasks without explicit programming. You will master Machine Learning concepts and techniques, including supervised and unsupervised learning, mathematical and heuristic aspects, and hands-on modeling to develop algorithms and prepare you for your role with advanced Machine Learning knowledge.

Key Learning Objectives

- ✓ Master the concepts of supervised and unsupervised learning, recommendation engine, and time series modeling
- ✓ Gain practical mastery over principles, algorithms, and applications of Machine Learning through a hands-on approach that includes working on four major end-to-end projects and 25+ hands-on exercises
- ✓ Acquire thorough knowledge of the statistical and heuristic aspects of Machine Learning
- ✓ Implement models such as support vector machines, kernel SVM, naive Bayes, decision tree classifier, random forest classifier, logistic regression, K-means clustering and more in Python
- ✓ Validate Machine Learning models and decode various accuracy metrics. Improve the final models using another set of optimization algorithms, which include Boosting & Bagging techniques
- ✓ Comprehend the theoretical concepts and how they relate to the practical aspects of Machine Learning

Course curriculum

- ✓ Lesson 01 - Introduction to Artificial Intelligence and Machine Learning
- ✓ Lesson 02 - Data Wrangling and Manipulation
- ✓ Lesson 03 - Supervised Learning
- ✓ Lesson 04 - Feature Engineering
- ✓ Lesson 05 - Supervised Learning Classification
- ✓ Lesson 06 - Unsupervised Learning
- ✓ Lesson 07 - Time Series Modeling
- ✓ Lesson 08 - Ensemble Learning
- ✓ Lesson 09 - Recommender Systems
- ✓ Lesson 10 - Text Mining

Tableau

This Tableau course helps you understand how to build visualizations, organize data, and design charts and dashboards to empower more meaningful business decisions. You'll be exposed to the concepts of Data Visualization, different combo charts, and stories, working with filters, parameters, and sets, and building interactive dashboards.

Key Learning Objectives

- ✓ Become an expert on visualization techniques such as heat map, treemap, waterfall, Pareto
- ✓ Understand metadata and its usage
- ✓ Work with Filter, Parameters, and Sets
- ✓ Master special field types and Tableau-generated fields and the process of creating and using parameters
- ✓ Learn how to build charts, interactive dashboards, story interfaces, and how to share your work
- ✓ Master the concepts of data blending, create data extracts and organize and format data
- ✓ Master arithmetic, logical, table, and LOD calculations

Course curriculum

- ✓ Lesson 01 - Getting Started with Tableau
- ✓ Lesson 02 - Core Tableau in Topics
- ✓ Lesson 03 - Creating Charts in Tableau
- ✓ Lesson 04 - Working with Metadata
- ✓ Lesson 05 - Filters in Tableau
- ✓ Lesson 06 - Applying Analytics to the worksheet
- ✓ Lesson 07 - Dashboard in Tableau
- ✓ Lesson 08 - Modifications to Data Connections
- ✓ Lesson 09 - Introduction to Level of Details in Tableau (LODS)

Natural Language Processing

This Natural Language Processing course will give you a detailed look at the science behind applying Machine Learning algorithms to process large amounts of natural language data. You will learn the concepts of statistical machine translation and neural models, deep semantic similarity model (DSSM), neural knowledge base embedding, deep reinforcement learning technique, neural models applied in image captioning, and visual question answering using Python's Natural Language Toolkit (NLTK).

Key Learning Objectives

- ✓ Apply Deep Learning models to solve machine translation and conversation problems
- ✓ Implement deep structured semantic models (DSSM) to retrieve information
- ✓ Understand deep reinforcement learning techniques applied in Natural Language Processing
- ✓ Use neural models applied in image captioning and visual question answering

Course curriculum

- ✓ Lesson 01 - Introduction to Natural Language Processing
- ✓ Lesson 02 - Feature Engineering on Text Data
- ✓ Lesson 03 - Natural Language Understanding Techniques
- ✓ Lesson 04 - Natural Language Generation
- ✓ Lesson 05 - Natural Language Processing Libraries
- ✓ Lesson 06 - Natural Language Processing with Machine Learning and Deep Learning
- ✓ Lesson 07 - Speech Recognition Technique

Data Science Capstone Project

This Data Science Capstone project will give you an opportunity to implement the skills you learned throughout this Program. Through dedicated mentoring sessions, you'll learn how to solve a real-world, industry-aligned Data Science problem, from data processing and model building to reporting your business results and insights. The project is the final step in the learning path and will enable you to showcase your expertise in Data Science to future employers.

Key Learning Objectives

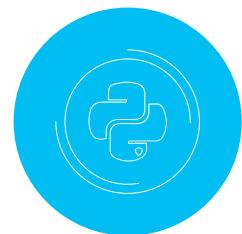
Simplilearn's online Data Science Capstone course will bring you through the Data Science decision cycle, including data processing, building a model and representing results. The project milestones are:

- ✓ **Data Processing** - In this step, you will apply various data processing techniques to make raw data meaningful.
- ✓ **Model Building** - You will leverage techniques such as regression and decision trees to build Machine Learning models that enable accurate and intelligent predictions. You may explore Python, R, or SAS to develop your model. You will follow the complete model-building exercise from data split to test and validate data using the k-fold cross-validation process.
- ✓ **Model Fine-tuning** - You will apply various techniques to improve the accuracy of your model and select the champion model that provides the best accuracy.
- ✓ **Dashboarding and Representing Results** - As the final step, you will be required to export your results into a dashboard with meaningful insights using Tableau.

Elective Course

IBM Watson for Chatbots

This course provides a practical introduction on how to build a chatbot with Watson Assistant without writing any code and then deploy your chatbot to a real website in less than five minutes. It will teach you to plan, build, test, analyze, and deploy your first chatbot.

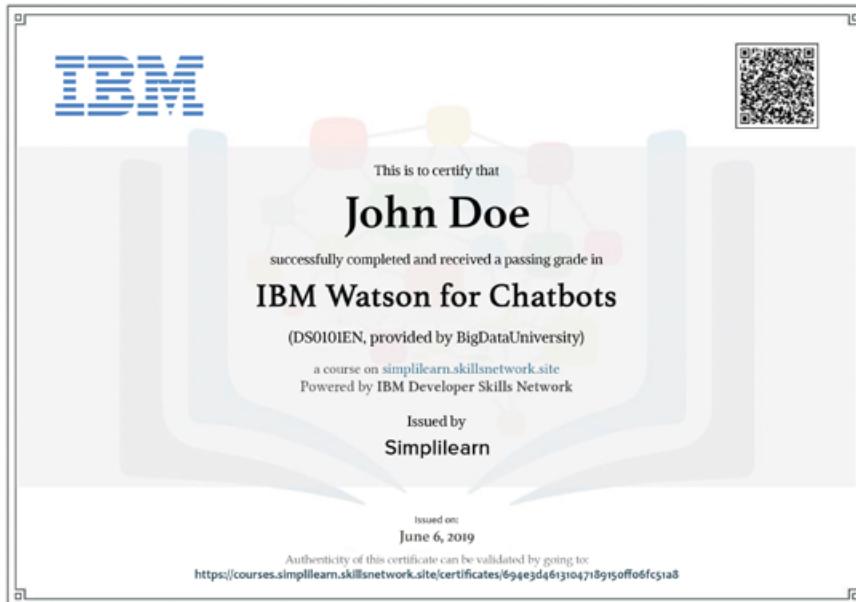


Machine Learning with R

In this course, you will learn how to write R code, learn about R's data structures, and create your own functions. With the knowledge gained, you will be ready to undertake your first very own data analysis. You'll further learn about supervised versus unsupervised learning, look into how statistical modeling relates to Machine Learning, and do a comparison of each using R.



Certification



Upon completion of this Post Graduate Program in Data Science in partnership with University, you will receive the Post Graduate certification from Purdue University and IBM. You will also receive certificates from Simplilearn for the Data Science courses in the learning path. These certificates will testify to your skills as an expert in Data Science.

Advisory Board Members



Dr. Christopher G. Brinton

[Assistant Professor of Electrical and Computer Engineering at Purdue University](#)

Dr. Christopher G. Brinton, assistant professor at Purdue University and co-founder of machine intelligence analytics company Zoomi, Inc., specializes in research on AI and networks. He has built award-winning social learning network algorithms and network optimization models.



Patrick J. Wolfe

[Frederick L. Hovde Dean of Science at Purdue University](#)

Patrick J. Wolfe, an award-winning researcher in the mathematical foundations of data science, is the Frederick L. Hovde Dean of Science at Purdue University and named the 2018 Distinguished Lecturer in Data Science by the IEEE. He provides expert advice on applications of data science.



Ronald Van Loon

Big Data Expert, Director - Adversitemet

Named by Onalytica as one of the three most influential people in Big Data, Ronald van Loon is an author for a number of leading Big Data and Data Science websites, including Datafloq, Data Science Central, and The Guardian. He is also a renowned speaker at industry events.



USA

Simplilearn Americas, Inc.
201 Spear Street, Suite 1100, San Francisco, CA 94105
United States
Phone No: +1-844-532-7688

INDIA

Simplilearn Solutions Pvt Ltd.
53/1 C, Manoj Arcade, 24th Main, HSR Layout
2nd Sector, HSR Layout
Bangalore - 560102
Call us at: 1800-212-7688

www.simplilearn.com

Disclaimer: All programs are offered on a non-credit basis and are not transferable to a degree.