

Data Scientist Implementer

Pratul Goyal

Email: pratul.goyal111@gmail.com

LinkedIn: <https://www.linkedin.com/in/pratul-goyal-7a7229119/>

Blogs: www.datasciencejourneyblogs.com

Contact: +916395353408

Objective

Data Scientist Implementer with master's in "Data Science and Business Analytics" from Indian Institute of Management with 6+ years of experience. Strong problem-solving and analytical skills. Advanced programming proficiency. Certified in Data Analysis and Machine Learning with a total of 6.5 years of experience.

Skills

- Tool expertise **R, Python, Excel, and Tableau.**
- Very strong background in Data Engineering, Machine Learning, Deep learning, Programming.
- Highly proficient in Classification, Clustering, Reinforcement Learning, Dimensionality Reduction (LDA, PCA, Kernel PCA) and Regression Algorithms.
- Strongly skilled in python libraries like TensorFlow, Matplotlib, Sklearn, NumPy, Pandas, PySpark.
- Strongly skilled in techniques like XGBoost, AdaBoost and K-Fold Cross Validation.
- Strongly skilled in R libraries like dplyr, e1071, tidyverse, Plotly and ggplot2.
- Skilled in handling SQL queries.
- Proficient knowledge of Excel: VLOOKUP, HLOOKUP, Pivoting, Nested Excel formulas.
- Skilled in Programming, Data Engineering, ANN, CNN, RNN, Auto-encoders.
- Exploratory data analysis, Data preprocessing, model evaluation.

Professional Certifications

<https://tinyurl.com/y64noado> : Link of all certifications

CERTIFICATIONS COURSERA

UNIVERSITY OF MICHIGAN SPECIALIZATION

- Introduction to Data Science in Python
- Applied Plotting, Charting & Data Representation in Python
- Applied Machine Learning in Python
- Applied Text Mining in Python
- Applied Social Network Analysis in Python

IBM DATA SCIENCE PROFESSIONAL CERTIFICATE

- What is Data Science?
- Tools for Data Science
- Data Science Methodology
- Python for Data Science and AI
- Databases and SQL for Data Science
- Data Analysis with Python
- Data Visualization with Python

- Machine Learning with Python
- Applied Data Science Capstone

IMPERIAL COLLEGE LONDON SPECIALIZATION

- Mathematics for Machine Learning: Linear Algebra
- Mathematics for Machine Learning: Multivariate Calculus
- Mathematics for Machine Learning: PCA

DEEPLARNING.AI SPECIALIZATION

- Neural Networks and Deep Learning
- Improving Deep Neural Networks: Hyperparameter tuning, Regularization and Optimization
- Structuring Machine Learning Projects
- Convolutional Neural Networks
- AI For Everyone

WHARTON ONLINE

- Operations Analytics
- People Analytics

INSEAD

- Introduction to Blockchain
- Transacting on the Blockchain

UDEMY

- Machine Learning A-Z Hands-On Python and R in Data Science
- Introduction to R
- Tableau 10 A-Z: Hands-On Tableau Training for Data Science!
- Data Science and Machine Learning with Python - Hands-On
- The Complete SQL Bootcamp

SIMPLILEARN

- Data Science with R
- Data Science with Python

Work History and Experience

SIMPLILEARN / DATA SCIENCE IMPLEMENTER / Bengaluru, India

03.2020-Current

- Conducted more than 750 hours of implementation on Data Science in R and Python.
- Proficient in modules on Machine Learning covering topics on Regression, Supervised and Unsupervised Algorithms.
- Proficient in Statistical testing viz. ANOVA, CHI-SQUARE, T-Test, Z-Test, Hypothesis testing with One tailed and Two tailed Test.
- Hands on expertise on Python, with proficiency on modules on Matplotlib, Pandas, Sklearn, PySpark, NumPy and Nltk.
- Hand on expertise on R, with proficiency on modules on dplyr, caret, e1071 and ggplot2, tidyverse, lubridate and Plotly.
- Proficient knowledge of Data structures on R and Python.
- Proficient in Web Scraping with BeautifulSoup Python Library.
- Hands on at 7 R and 4 Python Capstone Projects offered by Simplilearn.
- Proficient knowledge of handling SQL, and SQL queries in python using the PySpark Library.

ASSISTANT PROFESSOR / GRAPHIC ERA HILL UNIVERSITY / Dehradun, India 09.2018 – VISITING PROFESSOR

- Involved in teaching Business Analytics, Business Statistics, Ecommerce, Information System and Digital Marketing.
- Guided students on Data science research papers on International journals like Scopus, Sci, Taylor, and Francis Journals.
- Involved with Scholastic Seed Inc Journal, in Analytics editorial board.
- Principal Blogger at DataScienceJourney.com the analytics blogging website.
- Currently writing research paper on the application of MapReduce on Adjectival Bipartite Graph in this paper we have proposed an algorithmic implementation of an ASBG data visualization technique. This visualization technique is known as the adjectival bipartite sentiment graph for decision-making problems.

DATA SCIENCE TRAINER / THE PRINCETON REVIEW / Bengaluru, India 02.2018 – 08.2018

- Trained students for Data science concepts for Data science prep.

SUBJECT MATTER EXPERT / BYJU'S - THE LEARNING APP / Noida, India 03.2017 – 01.2018

- Mathematics Content Developer using Machine Learning Algorithms.
- Based on the response rate of Mathematics questions we were able to predict the response time per questions

MACHINE LEARNING TRAINER / THE PRINCETON REVIEW / Gurgaon, India 03.2016 – 03.2017

- Trained students for Machine Learning Concepts.
- Specialized in Python and Machine Learning.

DATA ANALYST / SVM INFOTECH / Noida, India 08.2013 – 02.2016

- Participated in requirements meetings and data mapping sessions to understand business needs.
- Performed R&D on available data to come up with a new and enhanced analysis of data.

Education

INDIAN INSTITUTE OF MANAGEMENT, KASHIPUR
EXECUTIVE MBA | 2015-2017

PANTNAGAR UNIVERSITY, PANTNAGAR
BTECH BIOTECHNOLOGY| 2009-2013

Project Experience

- **Customer Service Requests Analysis:** NYC 311's mission is to provide the public with quick and easy access to all New York City government services and information while offering the best customer service. Each day, NYC311 receives thousands of requests related to several hundred types of non-emergency services, including noise complaints, plumbing issues, and illegally parked cars. These requests are received by NYC311 and forwarded to the relevant agencies such as the police, buildings, or transportation. The agency responds to the request, addresses it, and then closes it.
- **Movielens Case Study:** The GroupLens Research Project is a research group in the Department of Computer Science and Engineering at the University of Minnesota.

Members of the GroupLens Research Project are involved in many research projects related to the fields of information filtering, collaborative filtering, and recommender systems. The project is led by professors John Riedl and Joseph Konstan. The project began to explore automated collaborative filtering in 1992 but is most well-known for its worldwide trial of an automated collaborative filtering system for Usenet news in 1996. Since then the project has expanded its scope to research overall information by filtering solutions, integrating into content-based methods, as well as, improving current collaborative filtering technology.

- **Comcast Telecom Consumer Complaints:** Comcast is an American global telecommunication company. The firm has been providing terrible customer service. They continue to fall short despite repeated promises to improve. Only last month (October 2016) the authority fined them a \$2.3 million, after receiving over 1000 consumer complaints. The existing database will serve as a repository of public customer complaints filed against Comcast. It will help to pin down what is wrong with Comcast's customer service.
- **Retail Analysis with Walmart Data:** One of the leading retail stores in the US, Walmart, would like to predict the sales and demand accurately. There are certain events and holidays which impact sales on each day. There are sales data available for 45 stores of Walmart. The business is facing a challenge due to unforeseen demands and runs out of stock sometimes, due to the inappropriate machine learning algorithm. An ideal ML algorithm will predict demand accurately and ingest factors like economic conditions including CPI, Unemployment Index, etc. Walmart runs several promotional markdown events throughout the year. These markdowns precede prominent holidays, the four largest of all, which are the Super Bowl, Labour Day, Thanksgiving, and Christmas. The weeks including these holidays are weighted five times higher in the evaluation than non-holiday weeks. Part of the challenge presented by this competition is modelling the effects of markdowns on these holiday weeks in the absence of complete/ideal historical data. Historical sales data for 45 Walmart stores located in different regions are available.
- **Insurance factors identification:** The data gives the details of third party motor insurance claims in Sweden for the year 1977. In Sweden, all motor insurance companies apply identical risk arguments to classify customers, and thus their portfolios and their claims statistics can be combined. The data were compiled by a Swedish Committee on the Analysis of Risk Premium in Motor Insurance. The Committee was asked to look into the problem of analysing the real influence on the claims of the risk arguments and to compare this structure with the actual tariff.
- **College Admission:** Every year thousands of applications are being submitted by international students for admission in colleges of the USA. It becomes an iterative task for the Education Department to know the total number of applications received and then compare that data with the total number of applications successfully accepted and visas processed. Hence to make the entire process easy, the education department in the US analyse the factors that influence the admission of a student into colleges. The objective of this exercise is to analyse the same.
- **High value customers identification for an E-Commerce company:** A UK-based online retail store has captured the sales data for different products for the period of one year (Nov 2016 to Dec 2017). The organization sells gifts primarily on the online platform. The customers who make a purchase consume directly for themselves. There are small businesses that buy in bulk and sell to other customers through the retail outlet channel.
- **Healthcare cost analysis:** A nationwide survey of hospital costs conducted by the US Agency for Healthcare consists of hospital records of inpatient samples. The given data is restricted to the city of Wisconsin and relates to patients in the age group 0-17 years. The agency wants to analyse the data to research on healthcare costs and their utilization.

Research Experience

MBA Research | Indian Institute of Management

- Variables that affect the Education System in India: Project is about measuring the efficiency of the Education system of India using DEA (Data Envelopment Analysis) and Regression techniques on Python.
- Optimization of Logistic for "BYJU'S ": This was the transshipment problem based on extra usage of the transportation cost by BYJU'S.

BTECH Research | Pantnagar University

- Enhanced Blastogenic Response in B and T Lymphocytes of Chicken Due to in vitro Exposure of Catharanthus roseus. • Created 3D-Molecular image of protein using Python Libraries Matplotlib and Seaborn.

Achievements

- Received the certificate of appreciation for outstanding performance as a Data Science Instructor by Princeton Review.
- Won National Level Paper Presentation at College Level.
- Cracked IIT-JEE examination with All India rank of 25000 in the year 2009.
- Received Outstanding letter of recommendation by Simplilearn for best training.
- Received letter of appreciation by BYJU'S The Learning App for outstanding work.
- Guided more than 1000+ working professionals being a part of Corporate Training.

Declaration

I hereby declare that all the information furnished above is accurate to the best of my knowledge and belief.

PRATUL GOYAL