

DataScientist 101

Complied by Nauman Jaffar

Anatomy of Modern Data Scientist

MODERN DATA SCIENTIST


Data Scientist, the sexiest job of the 21st century, requires a mixture of multidisciplinary skills ranging from an intersection of mathematics, statistics, computer science, communication and business. Finding a data scientist is hard. Finding people who understand who a data scientist is, is equally hard. So here is a little cheat sheet on who the modern data scientist really is.

MATH & STATISTICS

- ☆ Machine learning
- ☆ Statistical modeling
- ☆ Experiment design
- ☆ Bayesian inference
- ☆ Supervised learning: decision trees, random forests, logistic regression
- ☆ Unsupervised learning: clustering, dimensionality reduction
- ☆ Optimization: gradient descent and variants

PROGRAMMING & DATABASE

- ☆ Computer science fundamentals
- ☆ Scripting language e.g. Python
- ☆ Statistical computing packages, e.g., R
- ☆ Databases: SQL and NoSQL
- ☆ Relational algebra
- ☆ Parallel databases and parallel query processing
- ☆ MapReduce concepts
- ☆ Hadoop and Hive/Pig
- ☆ Custom reducers
- ☆ Experience with xaaS like AWS



DOMAIN KNOWLEDGE & SOFT SKILLS

- ☆ Passionate about the business
- ☆ Curious about data
- ☆ Influence without authority
- ☆ Hacker mindset
- ☆ Problem solver
- ☆ Strategic, proactive, creative, innovative and collaborative

COMMUNICATION & VISUALIZATION

- ☆ Able to engage with senior management
- ☆ Story telling skills
- ☆ Translate data-driven insights into decisions and actions
- ☆ Visual art design
- ☆ R packages like ggplot or lattice
- ☆ Knowledge of any of visualization tools e.g. Flare, D3.js, Tableau

MarketingDistillery.com is a group of practitioners in the area of e-commerce marketing. Our fields of expertise include: marketing strategy and optimization; customer tracking and on-site analytics; predictive analytics and econometrics; data warehousing and big data systems; marketing channel insights in Paid Search, SEO, Social, CRM and brand.


 (c) Krzysztof Zawadzki

Top Data Science, AI and Machine Learning Resources

- Article: [Difference between Machine Learning, Data Science, AI, Deep Learning...](#)
- Article: [What is Data Science? 24 Fundamental Articles Answering This Question](#)

- Article: [Hitchhiker's Guide to Data Science, Machine Learning, R, Python](#)
- Tutorial: [Data Science Cheat Sheet](#)
- Tutorial: [How to Become a Data Scientist - On Your Own](#)
- Tutorial: [Advanced Machine Learning with Basic Excel](#)
- Categories: [Data Science](#) - [Machine Learning](#) - [AI](#) - [IoT](#) - [Deep Learning](#)
- Tools: [Hadoop](#) - [DataViZ](#) - [Python](#) - [R](#) - [SQL](#) - [Excel](#)
- Techniques: [Clustering](#) - [Regression](#) - [SVM](#) - [Neural Nets](#) - [Ensembles](#) - [Decision Trees](#)
- Links: [Cheat Sheets](#) - [Books](#) - [Events](#) - [Webinars](#) - [Tutorials](#) - [Training](#) - [News](#) - [Jobs](#)
- Links: [Announcements](#) - [Salary Surveys](#) - [Data Sets](#) - [Certification](#) - [RSS Feeds](#) - [About Us](#)
- Newsletter: [Sign-up](#) - [Past Editions](#) - [Members-Only Section](#) - [Content Search](#) - [For Bloggers](#)

Machine Learning Overview

<https://youtu.be/o43NOKECjAw>