

My Best Data Science, Machine Learning and Related Articles

Posted by [Vincent Granville](#) on December 13, 2016 at 8:30am

<http://www.datasciencecentral.com/profiles/blogs/my-data-science-machine-learning-and-related-articles>

Here I list my most interesting contributions published on Data Science Central. My plan is to categorize and aggregate this content to produce a few self-published books. The material below will always be available for free (from this webpage), but the books won't, or if they are, they will be free [for members only](#). So you might want to bookmark this page.

I also have written a number of academic papers, [you can find some of them here](#).



My home office, where I write most of my DSC articles

The articles below are listed in reverse chronological order. This is a work in progress: I am still adding older entries. So check back again in a few weeks!

1. Core Articles

- [Data Science and Machine Learning Without Mathematics](#)
- [Advanced Machine Learning with Basic Excel](#)
- [State-of-the-Art Machine Learning Automation with HDT](#)
- [Tutorial: Neutralizing Outliers in Any Dimension](#)
- [The Fundamental Statistics Theorem Revisited](#)
- [Designing better algorithms: 5 case studies](#)
- [Life Cycle of Data Science Projects](#)
- [40 Techniques Used by Data Scientists](#)
- [Variance, Clustering, and Density Estimation Revisited](#)
- [Biased vs Unbiased: Debunking Statistical Myths](#)
- [The Death of the Statistical Tests of Hypotheses](#)
- [4 Easy Steps to Structure Highly Unstructured Big Data, via Automat...](#)
- [Architecture of Data Science Projects](#)
- [24 Uses of Statistical Modeling \(Part II\) | \(Part I\)](#)
- [The ABCD's of Business Optimization](#)
- [What you won't learn in stats classes](#)
- [21 data science systems used by Amazon to operate its business](#)
- [The best kept secret about linear and logistic regression](#)

- [Black-box Confidence Intervals: Excel and Perl Implementation](#)
- [Jackknife and linear regression in Excel: implementation and compar...](#)
- [Jackknife logistic and linear regression for clustering and predict...](#)

2. Blog Posts About Data Science

- [How to Detect if Numbers are Random or Not](#)
- [How to Solve the New \\$1 Million Kaggle Problem - Home Value Estimates](#)
- [Which classifier has the best performance?](#)
- [Introduction to Principal Component Analysis](#)
- [Detecting Fake News, Fake Reviews, Fake Accounts, Fake Pictures](#)
- [How and Why: Decorrelate Time Series](#)
- [Distribution of Arrival Times of Extreme Events](#)
- [10 Data Science, Machine Learning and IoT Predictions for 2017](#)
- [Modern Computational Advertising on Social Networks: The Basics](#)
- [Massive Internet Attack Floods the World with Fake Data](#)
- [Building an Algorithm to Break Strong Encryption](#)
- [Why so many Machine Learning Implementations Fail?](#)
- [MIT Algorithm Predicts Rogue Waves in Real Time to Save Lives](#)
- [What statisticians think about data scientists](#)
- [Why Zipf's law explains so many big data and physics phenomenons](#)
- [22 tips for better data science](#)

3. Other Blog Posts

- [Yet Another Interesting Math Problem - The Collatz Conjecture](#)
- [The Future of Scientific Publishing](#)
- [Simple Proof of the Prime Number Theorem](#)
- [Seasons in Binary Star Planetary Systems](#)
- [Factoring Massive Numbers: Machine Learning Approach](#)
- [88 percent of all integers have a factor under 100](#)
- [Math Challenge: Computing the Average Rotational Speed of Earth](#)
- [There Are Lies, Damn Lies, And Journalism](#)
- [Representation of Numbers as Infinite Products](#)
- [A Beautiful Probability Theorem](#)
- [Mars Craters: An Interesting Stochastic Geometry Problem](#)
- [The art and \(data\) science of leveraging economic bubbles](#)
- [Data Scientist Breaks State Monopoly on Lotteries](#)
- [Which jobs will AI \(Artificial Intelligence\) kill?](#)

4. Guides and References

- [Difference between Machine Learning, Data Science, AI, Deep Learnin...](#)
- [Data Science Cheat Sheet](#)
- [Data Science 2.0. \(About automated data science; in progress\)](#)
- [Developing Analytic Talent - Becoming a Data Scientist. Wiley, 2014.](#)
- [8 Deep Data Science Articles](#)
- [Hitchhiker's Guide to Data Science, Machine Learning, R, Python](#)
- [Deep Learning: Definition, Resources, Comparison with Machine Learning](#)
- [What is Data Science? 24 Fundamental Articles Answering This Question](#)
- [Answers to dozens of data science job interview questions](#)
- [15 Most Controversial Data Science Articles](#)
- [Challenges of the week](#)
- [Fascinating Facts and Conjectures about Primes and Other Special Nu...](#)
- [Three Original Math and Proba Challenges, with Tutorial](#)

5. Repositories

- [14 Timeless Reference Books](#)
- [20 Big Data Repositories You Should Check Out](#)
- [A Plethora of Data Set Repositories](#)
- [Four Great Pictures Illustrating Machine Learning Concepts](#)
- [15 Deep Learning Tutorials](#)
- [11 Great Hadoop, Spark and Map-Reduce Articles](#)
- [16 Great IoT Articles Published in 2016](#)
- [24 Core Articles About Data Science](#)
- [14 Great Articles and Tutorials on Clustering](#)
- [13 Great Articles and Tutorials about Correlation](#)
- [26 Great Articles and Tutorials about Regression Analysis](#)
- [10 Articles and Tutorials about Outliers](#)
- [21 Great Articles and Tutorials on Time Series](#)
- [15 Amazing Infographics and Other Visual Tutorials](#)
- [Misuses of Statistics: Examples and Solutions](#)
- [20 Cheat Sheets: Python, ML, Data Science, R, and More](#)

Follow us on Twitter: [@DataScienceCtrl](#) | [@AnalyticBridge](#)