College of Computing and Informatics, Drexel University
INFO 103 Introduction to Data Science - 3 credit hours
Online Section

Term: Spring 2018
Instructor: Jeremy Leipzig
Contact: jnl47@drexel.edu

Prerequisites
None

Curriculum role
This course is a core course for the BSDS program, a required course for the BSDS Minor, and a course in the Department of Information Science's common freshman year.

Required texts
There is no required textbook. This course is self-contained and open source; required readings include the course notes and any linked material labeled as "required." All materials may be found linked from the course notes (leipzig.github.io/Intro2DS/) and from the index at the bottom of the syllabus.

Description
A first course in data science. Introduces data science as a field, describes the roles and services that various members of the community play and the life cycle of data science projects. Provides an overview of common types of data, where they come from, and the challenges that practitioners face in the modern world of "Big Data." Provides an introduction to the interdisciplinary mixture of skills that the practice requires.

This course was designed by Dr. Jake Williams – it was originally INFO 240.

Rationale
Individuals from a wide range of disciplines work with or alongside technology and will, at some point or another in their careers, encounter the effects that the increasing presence of electronic data is making on how business operates and decisions are made. This survey course is intended to provide students with base-level knowledge on the emerging field of data science. Students may then go on to more advanced study in data science, or carry general knowledge of it forward into their own disciplines.

Learning outcomes
Upon successful completion of this course, students will be able to:

- discuss the main concepts of data science;
- illustrate the types of problems that data analytics can approach;
- describe a variety of data types and storage formats;
- explain challenges with processing Big Data efficiently and effectively;
- identify the skills and capabilities needed by data scientists;
- and interpret the key features of a complete data science project.

Discussion
The online section of the course will involve the participation in the Blackboard discussion board. Topics will be derived from the readings

Demos
There will be 8 programming demonstrations that will represent course concepts concretely in the context of the Python programming language. These demonstrations will be presented in class and are be read and interpreted, only. While programming is not a curricular requirement of this course, student understanding of the concepts and challenges illustrated in the programming demos will be assessed in the homework.

Homework
There will be two homework assignments to be **completed individually**. These will cover course readings and demos, will have two-week lifetimes, and are to be submitted electronically via Blackboard Learn **no later than 11:59pm** on the days due, with **names, IDs, and the course and assignment numbers** recorded in submission headers.

If you are dissatisfied with a grade or point deduction, you can request re-marking. Re-marking requests must be done through written (paper or email) descriptions of why you think the grade is in error. If you wish to appeal a course grade, school policies apply.

Mid-term exam
There will be one mid-term exam on week 6, conducted in Blackboard. There is no final exam.

Group project
There will be a project to be completed by groups of size 3 (or 4, depending on class size), whose milestones will consist of a written description, due on the Friday of week 8, and a presentation to be conducted as a video or a voiceover slide deck in week 10. There is no final exam.

Grading
Course grades will be based on Blackboard discussion participation (**25%**), two homework assignments (**25%**), one exam (**20%**), and a group project's written description (**15%**) and presentation (**15%**):

Milestone	Weight	Material covered	Due
Discussion	25%	Readings	Weeks 1–5 & 7–9
Homework	25%	Readings, demos	Weeks 3 & 5
Mid-term	20%	Readings, discussions	Week 6
Description	15%	Group-identified topic	Week 8
Presentation	15%	Group-identified topic	Weeks 10+

Grades are assessed according to content (correctness) and form (presentation). Conversion from points to letters is performed as follows:

97 – 100	A+	77 – 80	C+
94 – 97	A	74 – 77	C

90 – 94	A-	70 – 74	C-
87 – 90	B+	67 – 70	D+
84 – 87	B	64 – 67	D
80 – 84	B-	60 – 64	D-
		0 – 60	F

Schedule

The schedule is tentative and will cover one or more chapters per week, with variation in accordance with differences in student preparation, class discussion, and chapter length. The table below shows the anticipated schedule.

Week	Readings	Milestones
1	Ch. 0: 1; Ch. 1: 1–10; Ch. 2: 1–4	Discussion 1, Demo 1
2	Ch. 3: 1–5; Ch. 4: 1–5	Discussion 2, Demo 2
3	Ch. 5: 1–4; Ch. 6: 1–6	Demo 3, Homework 1 Due
4	Ch. 7: 1–6; Ch. 8: 1–4	Discussion 4, Demo 4
5	Ch. 8: 5–8; Ch. 9: 1–3;	Homework 2
6	Ch. 10: 1–7;	Exam 1, Demo 5, Projects ideas
7	Ch. 11: 1–7;	Discussion 5, Demo 6
8	Ch. 11: 8–15;	Demo 7, Proposal due
9	Ch. 12: 1–4; Ch. 13: 1–4	Discussion 6, Demo 8
10	Ch. 14: 1–9	Reports & Presentations Due Last Day of Class

Disabilities

Students requesting accommodations due to a disability at Drexel University need to present a current Accommodation Verification Letter (AVL) to faculty before accommodations can be made. AVL's are issued by the Office of Disability Resources (ODR). For additional information, visit the ODR website at http://www.drexel.edu/oed/disabilityResources, or contact the Office for more information: 215-895-1401 (V), or disability@drexel.edu.

Withdrawal of the Course

For dropping or withdrawing from the course, please refer to the university policies at:

- http://www.drexel.edu/provost/policies/course-add-drop/
- http://www.drexel.edu/provost/policies/course-withdrawal/

Class Lecture Recording
Lectures may be rebroadcast for educational purposes only.

Incomplete Policy
Incomplete grades are contingent upon instructor approval and will only be considered in extenuating circumstances beyond the student's control. The instructor is under no obligation to offer an incomplete grade. At least 80% of the graded coursework must have already been completed in order for an incomplete grade to be considered (per the recommendation of the Provost's Office). An incomplete contract with an instructor-determined due date for delivery of the completed work must be completed by the student and the instructor. It can be found here: http://www.drexel.edu/provost/policies/pdf/forms/incomplete.pdf.

Support for Equality and Diversity
Drexel University strives to promote an environment of equality of opportunity and compliance with University policies and federal, state and local laws prohibiting discrimination based upon race, color, religion, gender (sex), marital status, pregnancy, national origin, age, disability and veteran status. Students, faculty, and staff with questions about or complaints concerning discrimination, harassment, and/or retaliation should contact the Office of Equality and Diversity at (215) 895-1403 or http://www.drexel.edu/oed/

Student Conduct and Community Standards
Drexel University expects that all students as well as student organizations will conduct themselves responsibly and in a manner that reflects favorably upon themselves and the University. Check out here: http://www.drexel.edu/studentlife/community_standards/overview/ for the university policies, rules, regulations, and standards of conduct.

Academic Honesty
The Drexel University Academic Honesty Rules and Procedures (as stated in the student handbook) will be adhered to strictly. Students who commit plagiarism or cheat on assignments may receive an F grade for both the assignment and the course.

In order to avoid plagiarizing material, observe the following:

- If you work on an assignment with another student or a group of students, be certain that your final, individual paper is your own work or the work of your project group (for group projects) unless otherwise specified by the professor. While you might want to discuss the assignment with other students, you must, in your paper, express your own ideas in your own way.
- If you use printed or electronic resources in your papers, be sure to attribute the sources you have used. This can be done by quoting the material or by paraphrasing the material and, in either case, listing the source in an annotated bibliography. Use standard notation when citing references.

The college is requiring students to append a statement to deliverables (for example: papers, projects, exams) indicating that the work submitted is their own. Deliverables will not be graded without a separate certification page.

Please sign the following academic honesty statements by typing your name and date, and submit it through the Blackboard system as an individual assignment.

==
I certify that my work in this course will be entirely my own work. I will not quote the words of any other person from a printed source or a website without indicating what has been quoted and providing an appropriate citation. I will not submit my work in this course to satisfy the requirements of any other course.

Name/Signature_________________________
Date ___________________
==

Index

0. **Introduction**: Outline of course intentions, content, and expectations.

    1. Course syllabus;
    Introduction to Data Science; 4/3/2018
    leipzig.github.io/Intro2DS/syllabus.pdf

1. **Overview**: Broad description of data science, its history, and practitioners.

    1. **Data Science and Data Scientists: What's in a Name?**;
    Todd Saunders; Information Management; 9/11/2013
    http://www.cbigconsulting.com/wp-content/uploads/2014/03/data-scientist.pdf
    2. **Data science has become a well established discipline, so what is it?**;
    Jonathan Sedar; The Sampler; 2/11/2015
    http://blog.applied.ai/what-is-data-science/
    3. **Interview: Senior Director of Data Science at Cloudera**;
    Josh Wills; The Data analytics handbook, pgs. 8--12; 2014
    https://s3.amazonaws.com/leada/handbook/Handbook_Pt1.pdf
    4. **Introduction: What Is Data Science?**;
    Cathy O'Neil & Rachel Schutt;
    Doing Data Science: Straight Talk From the Frontline, pgs 1--16; 2014
    http://cdn.oreillystatic.com/oreilly/booksamplers/9781449358655_sampler.pdf
    5. **The Data Science Venn Diagram; Drew Conway**;
    Zero Intelligence Agents; 9/30/2010
    http://drewconway.com/zia/2013/3/26/the-data-science-venn-diagram
    6. **The Fourth Bubble in the Data Science Venn Diagram: Social Sciences**;
    Michael Malak; Data Science Association; 2/28/2014
    http://datascienceassn.org/content/fourth-bubble-data-science-venn-diagram-social-sciences
    7. **Data Science Falls Into Many Roles**;
    Rawn Shah; Forbes; 10/6/2015
    http://www.forbes.com/sites/rawnshah/2015/10/06/data-science-falls-into-many-roles/#5d7ec2b335f8
    8. **A Very Short History Of Data Science**;
    Gil Press; Forbes; 5/28/2013
    http://www.forbes.com/sites/gilpress/2013/05/28/a-very-short-history-of-data-science/#5cbf3a6169fd
    9. **A History of Data Science (2. Discussion)**;
    Wikimedia Foundation: Wikibooks; Data Science: An Introduction, Sec. 1.2; 8/24/2016
    https://en.wikibooks.org/wiki/Data_Science:_An_Introduction/A_History_of_Data_Science#Discussion
    10. **The Internet of Things Is Far Bigger Than Anyone Realizes**;
    Daniel Burrus; Wired; 11/1/2014

https://www.wired.com/insights/2014/11/the-internet-of-things-bigger/

2. **Life cycle**: Essential components of a data science project and best practices.

1. **The Goal is Data Product: Now How Do We Get There?**;
Ryan Swanstrom; Data Science 101; 1/12/2015
http://101.datascience.community/tag/lifecycle/

2. **Cross Industry Standard Process for Data Mining**;
Wikimedia Foundation: Wikipedia; 4/29/2016
https://en.wikipedia.org/wiki/Cross_Industry_Standard_Process_for_Data_Mining

3. **The data science project lifecycle**;
Maloy Manna; Linked in Pulse; 1/8/2015
https://www.linkedin.com/pulse/data-science-project-lifecycle-maloy-manna-pmp-

4. **Data Science Workflow: Overview and Challenges**;
Philip Guo; Communications of the ACM Blog; 10/30/2013
http://cacm.acm.org/blogs/blog-cacm/169199-data-science-workflow-overview-and-challenges/fulltext

3. **Data types**: Discussion of types of data and their relationships.

1. **Bridging the Divide between Unstructured and Structured Data**;
Jim Harris; UC Berkeley Data Science Blog; 2/26/2014
https://datascience.berkeley.edu/structured-unstructured-data/

2. An introduction to text analytics
https://www.datasciencecentral.com/profiles/blogs/an-introduction-to-text-analytics

3. **Creating Web Designs: Raster Versus Vector Graphics**;
Office of Teaching and Learning, Utah Valley University; DGM 2740, Web Design
http://desource.uvu.edu/dgm/2740/IN/steinja/lessons/05/l05_06.html?m=1

4.**What is the difference between categorical, ordinal and interval variables?**;
Introduction to SAS. UCLA: Statistical Consulting Group. (accessed 12/21/2016)
http://www.ats.ucla.edu/stat/mult_pkg/whatstat/nominal_ordinal_interval.htm

5. **Metadata (1. History, 2. Definition, 3. Types, and 6. Use)**;
https://en.wikipedia.org/wiki/Metadata
Wikimedia Foundation: Wikipedia; 12/19/2016

6. Understanding metadata
https://groups.niso.org/apps/group_public/download.php/17446/Understanding Metadata.pdf

7. Taxonomies and folksonomies
http://knowledgebird.com/2012/12/05/taxonomies-and-folksonomies/

8. The Semantic Web
https://www.ted.com/talks/tim_berners_lee_on_the_next_web

4. "**Big data**:" Characterizations of Big Data in the context of sources.

1. **A Personal Perspective on the Origin(s) and Development of "Big Data": The Phenomenon, the Term, and the Discipline**;
Francis X. Diebold; 9/26/2012
http://www.ssc.upenn.edu/~fdiebold/papers/paper112/Diebold_Big_Data.pdf

2. **3D Data Management: Controlling Data Volume, Velocity, and Variety**;
Doug Laney; Application Delivery Strategies; 2/6/2001
https://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf

3. **Big Data: Avoid 'Wanna V' Confusion**;
Seth Grimes; InformationWeek; 8/7/2013

http://www.informationweek.com/big-data/big-data-analytics/big-data-avoid-wanna-v-confusion/d/d-id/1111077?

4. **Update on the Twitter Archive at the Library of Congress; Library of Congress**; Whitepaper; January, 2013
jakerylandwilliams.github.io/Intro2DS/readings/twitter_report_2013jan.pdf

5. **Library of Congress' Twitter archive is a huge #FAIL**;
Nancy Scola; Politico; 7/11/2015
http://www.politico.com/story/2015/07/library-of-congress-twitter-archive-119698.html

5. **Technology**: Methods and technologies for handling Big Data.

1. **Embarrassingly parallel (0. Introduction, 1. Etymology)**;
Wikimedia Foundation: Wikipedia; 12/12/2016
https://en.wikipedia.org/wiki/Embarrassingly_parallel

2. **IO-Bound**
https://en.wikipedia.org/wiki/I/O_bound

3. **What is MapReduce?**;
IBM & Rafael Coss
https://www.youtube.com/watch?v=8wjvMyc01QY

4. **Conceptual Overview of Map-Reduce and Hadoop**;
Glenn Klockwood; Data-Intensive Computing; 10/9/2015
http://www.glennklockwood.com/data-intensive/hadoop/overview.html

5. **Apache Spark vs. MapReduce #WhiteboardWalkthrough**
MapR Data Technologies
https://www.youtube.com/watch?v=KzFe4T0PwQ8

6. Which Languages Should you Learn for Data Science?
https://medium.freecodecamp.org/which-languages-should-you-learn-for-data-science-e806ba55a81f

6. **Acquisition**: Identifying sources of data and considerations for acquisition.

1. **The Internet's hidden science factory**;
Jenny Marder and Mike Fritz; PBS NewsHour; 2/11/2015
http://www.pbs.org/newshour/updates/inside-amazons-hidden-science-factory/

2. **Hello Barbie: Considering Potential Unforeseen Problems With A.I. Dolls and What Children Tell Them**;
Amanda Zink; bioethics.net blog; 9/28/2015
http://www.bioethics.net/2015/09/hello-barbie-considering-potential-unforeseen-problems-with-a-i-dolls-and-what-children-tell-them/

3. **Privacy and Information Sharing**;
Lee Rainie and Maeve Duggan; PewResearch Report; 1/14/2016
http://www.pewinternet.org/2016/01/14/privacy-and-information-sharing/

4. **I Don't Need No Stinking API: Web Scraping For Fun and Profit**;
Hartley Brody; Hartley Brody Blog; 12/8/2012
https://blog.hartleybrody.com/web-scraping/

5. **Web scraping: legal or illegal?**;
ScrapeSentry; Scraping Wiki (accessed 12/21/2016)
https://www.scrapesentry.com/scraping-wiki/web-scraping-legal-or-illegal/

6. **Open data (3. Arguments for and Against)**;
Wikimedia Foundation: Wikipedia; 12/4/2016
https://en.wikipedia.org/wiki/Open_data#Arguments_for_and_against_open_data

7. **Pre-processing**: Transforming, cleaning, and enriching data.

1. **In Big Data, Preparing the Data is Most of the Work**;
Sullexis LLC; Data Science Central; 1/13/2015
http://www.datasciencecentral.com/profiles/blogs/in-big-data-preparing-the-data-is-most-of-the-work
2. **Data wrangling**;
Wikimedia Foundation: Wikipedia; 12/13/2016
https://en.wikipedia.org/wiki/Data_wrangling
3. **For Big-Data Scientists, 'Janitor Work' Is Key Hurdle to Insights**;
Steve Lohr; The New York Times; 8/17/2014
https://www.nytimes.com/2014/08/18/technology/for-big-data-scientists-hurdle-to-insights-is-janitor-work.html?_r=1
4.Tidy Data - Journal of Statistical Software
https://www.jstatsoft.org/article/view/v059i10/v59i10.pdf
5. **How Computers Generate Random Numbers**;
Chris Hoffman; HowToGeek; 2/22/2014
http://www.howtogeek.com/183051/htg-explains-how-computers-generate-random-numbers/
6. **An Introduction to Feature Selection**;
Jason Brownlee; Machine Learning Mastery; 10/6/2014
http://machinelearningmastery.com/an-introduction-to-feature-selection/
7. **Camouflage from face detection**;
Adam Harvey; CVDazzle (accessed 12/21/2016)
https://cvdazzle.com

8. **Storage**: Methods of data storage with considerations for access.

1. **Where is the World's Data Being Stored?**;
Mozy; 2007
https://mozy.com/infographics/where-is-the-worlds-data-stored/
2. **When Will We Run Out of Space?**;
Mozy; 2011
https://mozy.com/infographics/when-will-we-run-out-of-space/
3. **The History of Digital Storage**;
Matt Silverman (article) Mike Vasilev (Infographic); 10/8/2011
http://mashable.com/2011/10/08/digital-storage-infographic/#bqOHGIN_iEqV
4. **What is Cloud Computing?**;
Eric Griffith; PC Magazine; 5/3/2016
http://www.pcmag.com/article2/0,2817,2372163,00.asp
5. **Cloud computing security (5. Data security)**;
Wikimedia foundation: Wikipedia; 12/7/2016
https://en.wikipedia.org/wiki/Cloud_computing_security#Data_security
6. **How Europe Protects Your Online Data Differently Than the U.S.**;
Mark Scott & Natasha Singer; The New York Times; 1/31/2016
http://www.nytimes.com/interactive/2016/01/29/technology/data-privacy-policy-us-europe.html
7. **What are relational databases?**;
HowStuffWorks.com; 4/23/2001
http://computer.howstuffworks.com/question599.htm
8. **NoSQL Database: New Era of Databases for Big data Analytics - Classification, Characteristics and Comparison**
https://arxiv.org/pdf/1307.0191.pdf

9. **Description**: Data summarization methods and exploratory analysis.

1. **Understand Your Problem and Get Better Results Using Exploratory Data Analysis**;
Jason Brownlee; Machine Learning Mastery; 9/21/2014
http://machinelearningmastery.com/understand-problem-get-better-results-using-exploratory-data-analysis/
2. **How To Analyze Data Using the Average**;
Kalid; betterexplained.com; 2008
https://betterexplained.com/articles/how-to-analyze-data-using-the-average/
3. **Causation vs Correlation**;
Rebecca Goldin; STATS.org; 8/19/2015
http://www.stats.org/causation-vs-correlation/

10. **Depiction**: Data visualization types and more exploratory analysis.

1. **Common Map Projections**;
Rachel Quist; GIS Lounge; 12/3/2011
https://www.gislounge.com/common-map-projections/
2. **Map Projections**;
Randall Munroe; xkcd; 2011
https://xkcd.com/977/
3. **Comparing Map Projections**;
Kai S. Chang; bl.ocks.org; 9/4/2016
https://bl.ocks.org/syntagmatic/ba569633d51ebec6ec6e
4. **Data Storytelling: The Essential Data Science Skill Everyone Needs**;
Brent Dykes; Forbes; 4/31/2016
http://www.forbes.com/sites/brentdykes/2016/03/31/data-storytelling-the-essential-data-science-skill-everyone-needs/#59adf720f0c8
5. **How to Tell a Story with Data**;
Jim Stikeleather; Harvard Business Review; 4/24/2013
https://hbr.org/2013/04/how-to-tell-a-story-with-data
6. **Storytelling and data: when beautiful metrics can't beat words**;
Scott Brinker; Chief Marketing Technologist Blog; 4/13/2016
http://chiefmartec.com/2016/04/storytelling-data-beautiful-metrics-cant-beat-words/
7. W**hy data storytelling is so important—and why we're so bad at it**;
Tom Davenport; Deloitte University Press; 1/22/2015
https://dupress.deloitte.com/dup-us-en/topics/analytics/data-driven-storytelling.html

11. **Modeling**: Mathematical and statistical frameworks and models for data analysis.

1. **Calculus (2.2 Differential calculus)**;
Wikimedia Foundation: Wikipedia; 12/18/2016
https://en.wikipedia.org/wiki/Calculus#Differential_calculus
2. **Calculus (2.4 Integral calculus)**;
Wikimedia Foundation: Wikipedia; 12/18/2016
https://en.wikipedia.org/wiki/Calculus#Integral_calculus
3. **Essence of linear algebra (0. preview)**;
Grant Sanderson; 3blue1brown; 8/4/2016
https://www.youtube.com/watch?v=kjBOesZCoqc&index=1&list=PLZHQObOWTQDPD3MizzM2xVFitgF8hE_ab
4. **Essence of linear algebra (1. Vectors, what even are they?)**;

Grant Sanderson; 3blue1brown; 8/4/2016
https://www.youtube.com/watch?v=fNk_zzaMoSs&index=2&t=12s&list=PLZHQObOWTQDPD3MizzM2xVFitgF8hE_ab
5. **Essence of linear algebra (2. Linear combinations, span, and basis vectors)**;
Grant Sanderson; 3blue1brown; 8/4/2016
https://www.youtube.com/watch?v=k7RM-ot2NWY&index=3&t=60s&list=PLZHQObOWTQDPD3MizzM2xVFitgF8hE_ab
6. **Essence of linear algebra (3. Linear transformations and matrices)**;
Grant Sanderson; 3blue1brown; 8/4/2016
https://www.youtube.com/watch?v=kYB8IZa5AuE&index=4&t=311s&list=PLZHQObOWTQDPD3MizzM2xVFitgF8hE_ab
7. **Calculus Is So Last Century**;
Tianhui Miachel Li & Allison Bishop; The Wall Street Journal; 4/4/2016
http://www.wsj.com/articles/calculus-is-so-last-century-1457132991
8. **A Brief Introduction to Probability & Statistics**;
Kalid; betterexplained.com; 2012
https://betterexplained.com/articles/a-brief-introduction-to-probability-statistics/
9. **This is the difference between statistics and data science**;
Justin Megahan; The Signal; 4/30/2016
https://blog.mixpanel.com/2016/03/30/this-is-the-difference-between-statistics-and-data-science/
10. **Science Isn't Broken (It's just a hell of a lot harder than we give it credit for)**;
Christie Aschwanden; fivethirtyeight.com; 8/19/2015
http://fivethirtyeight.com/features/science-isnt-broken/
11. **A Visual Introduction to Machine Learning**;
Stephanie Yee & Tony Chu; R2D3.us; 2015
http://www.r2d3.us/visual-intro-to-machine-learning-part-1/
12. **Difference between Machine Learning & Statistical Modeling**;
Tavish Srivastava; Analytics Vidhya; 7/1/2015
https://www.analyticsvidhya.com/blog/2015/07/difference-machine-learning-statistical-modeling/
13. **How do Data Scientists perform model selection? Is it different from Kaggle?**;
Sebastian Raschka; Machine Learning FAQ; 2013--2016
https://sebastianraschka.com/faq/docs/model-selection-in-datascience.html
14. **Gentle Introduction to the Bias-Variance Trade-Off in Machine Learning**;
Jason Brownlee; Machine Learning Mastery; 4/18/2016
http://machinelearningmastery.com/gentle-introduction-to-the-bias-variance-trade-off-in-machine-learning/
15. **The End of Theory: The Data Deluge Makes the Scientific Method Obsolete**;
Chris Anderson; Wired; 6/23/08
jakerylandwilliams.github.io/Intro2DS/readings/the-end-of-theory.pdf

12. **Classification**: Methods and evaluations for algorithmic discrimination.
1. **Simple guide to confusion matrix terminology**;
Kevin Markham; Data School; 4/25/2014
http://www.dataschool.io/simple-guide-to-confusion-matrix-terminology/
2. **Classification Accuracy is Not Enough: More Performance Measures You Can Use**;
Jason Brownlee; Machine Learning Mastery; 4/21/2014

http://machinelearningmastery.com/classification-accuracy-is-not-enough-more-performance-measures-you-can-use/

3. **ROC Area-Under-the-Curve Explained**;
Konstantin Tretyakov; Fours Years Remaining; 12/10/2011
http://fouryears.eu/2011/10/12/roc-area-under-the-curve-explained/

4. **When Algorithms Discriminate**;
Claire Cain Miller; The New York Times; 7/9/2015
http://www.nytimes.com/2015/07/10/upshot/when-algorithms-discriminate.html

5. Software 'no more accurate than untrained humans' at judging reoffending risk
Hannah Devlin; The Guardian; 01/17/2018
https://www.theguardian.com/us-news/2018/jan/17/software-no-more-accurate-than-untrained-humans-at-judging-reoffending-risk

13. **Design**: Optimizing system performance with thoughtful design.

1. **Design Thinking for Data Scientists**;
George Roumeliotis; Intuit; 2/22/2015
https://www.youtube.com/watch?v=bOJlAKKF3eY

2. **How Data and Design Can Work Together**;
Heiko Waechter; Huge; 5/10/2016
http://www.hugeinc.com/ideas/perspective/how-data-and-design-can-work-together

3. **The best interface is no interface**;
Golden Krishna; Cooper; 8/29/2012
https://www.cooper.com/journal/2012/08/the-best-interface-is-no-interface

4. **Natural user interface**;
Wikimedia Foundation: Wikipedia; 10/2/2016
https://en.wikipedia.org/wiki/Natural_user_interface

5. **Practical Data Science and the Tricky Business of A/B Testing**;
James Kobielus; Dataversity; 6/2/2014
http://www.dataversity.net/practical-data-science-tricky-business-ab-testing/

6. **Doing Data Science at Twitter: A reflection of my two year Journey so far. Sample size N = 1**;
Robert Chang; Medium; 6/20/2015
https://medium.com/@rchang/my-two-year-journey-as-a-data-scientist-at-twitter-f0c13298aee6 - .gc9l5kujz

14. **Summary**: A walkthrough of an exemplar data science project history.

1. **Web search engine (1. History)**;
Wikimedia Foundation: Wikipedia; 12/20/2016
https://en.wikipedia.org/wiki/Web_search_engine#History

2. **Inverted Index**;
Wikimedia Foundation: Wikipedia: 9/21/2016
https://en.wikipedia.org/wiki/Inverted_index

3. **Finding more high-quality sites in search**;
Amit Singhal & Matt Cutts; Official Google Blog; 2/24/2011
https://googleblog.blogspot.com/2011/02/finding-more-high-quality-sites-in.html

4. **Google Clamps Down on Content Factories**;
Ryan Singel; Wired; 2/25/2011
https://www.wired.com/2011/02/google-clamp-down-content-factories/

5. **Google: We're Working to Help Good Sites Caught by Spam Cleanup**;
Ryan Singel; Wired; 3/1/2011

https://www.wired.com/2011/03/google-spam-side-effects/
6. **Personalized Search for everyone**;
Bryan Horling & Matthew Kulick; Official Google Blog; 12/4/2009
https://googleblog.blogspot.com/2009/12/personalized-search-for-everyone.html
7. **Google Personalized Search (4. Reception)**;
Wikimedia Foundation: Wikipedia; 10/12/2016
https://en.wikipedia.org/wiki/Google_Personalized_Search#Reception
8. **How to Turn off Google's Personalized Search Results**;
Julia Angwin; The Wall Street Journal Blog; 9/4/2012
http://blogs.wsj.com/digits/2012/11/04/how-to-turn-off-googles-personalized-search-results/
9. **Filter Bubble**;
Wikimedia Foundation: Wikipedia; 12/17/2016
https://en.wikipedia.org/wiki/Filter_bubble