Course Links Reference Guide

Dave Valentine

Contents

Section #2 – Core Concepts

Lecture #8 - Computer Science - the `Train Wreck' Definition

- Computer Science Definition from Wikipedia, the free encyclopedia
  - https://en.wikipedia.org/w/index.php?title=Computer_science&oldid=740982864

Lecture #9 - What's Data / "I can see data everywhere!"

- ASCII Definition from Wikipedia, the free encyclopedia
  - https://simple.wikipedia.org/w/index.php?title=ASCII&oldid=5469821

- File:Tst.png / From Wikimedia Commons, the free media repository
  - https://commons.wikimedia.org/wiki/File:Tst.png

Lecture #11 - Computer Science - Definition Revisited & The Greatest "lie" ever SOLD....

- On the Media / ROBOTS! (and artificial intelligence) / April 17, 2014 Episode
  - https://www.wnycstudios.org/podcasts/otm/episodes/on-the-media-2014-04-18

Lecture #12 - What's big data?

- Apache Spark
  - http://spark.apache.org/

- Apache Hadoop
  - http://hadoop.apache.org/

Lecture #13 - What is Artificial Intelligence (AI)

- Edsger W. Dijkstra biography from Wikipedia, the free encyclopedia
  - https://en.wikipedia.org/wiki/Edsger_W._Dijkstra

- Artificial intelligence Definition from Wikipedia, the free encyclopedia
  - https://simple.wikipedia.org/w/index.php?title=Artificial_intelligence&oldid=5408961

- Tic-tac-toe game Definition from Wikipedia, the free encyclopedia
  - https://en.wikipedia.org/wiki/Tic-tac-toe

- Radiolab / Mapping Tic Tac Toe-dom / September 6, 2011
  - https://www.wnycstudios.org/podcasts/radiolab/articles/156900-three-row

- Lighthill report from Wikipedia, the free encyclopedia
  - https://en.wikipedia.org/wiki/Lighthill_report

Lecture #14 - What is Machine Learning? - Part 1 - The ideas

- Machine learning Definition from Wikipedia, the free encyclopedia
  - https://en.wikipedia.org/w/index.php?title=Machine_learning&oldid=741484832

- Netflix Prize description from Wikipedia, the free encyclopedia
  - https://en.wikipedia.org/w/index.php?title=Netflix_Prize&oldid=727682948

- A 195-page monograph by a top-1% Netflix Prize contestant
  - https://www.kdnuggets.com/2016/03/netflix-prize-analyzed-movie-ratings-recommender-systems.html

- Got Lost? – Story of Randy Bilyeu
  - https://www.msn.com/en-ca/news/weekendreads/a-treasure-hunter-went-missing-in-the-rocky-mountains-and-a-computer-algorithm-found-him-months-later/ar-AAil0Pz

- You don't know her, but Crystal knows you
  - https://money.cnn.com/2015/09/03/technology/startup-crystal/index.html

- Note to Self / This Is How Much the Internet Knows About You / May 26, 2015
  - https://www.wnycstudios.org/podcasts/notetoself/episodes/crystal-knows-email

Lecture #15 - What is data science?

- Machine Learning Isn't Data Science / by Nwokedi C. Idika / Oct 31, 2015
  - https://medium.com/@nwokedi/machine-learning-isn-t-data-science-67cc66867dbc#.tvrvg9re7

- The Data Science Venn Diagram by Drew Conway
  - http://drewconway.com/zia/2013/3/26/the-data-science-venn-diagram

- Battle of the Data Science Venn Diagrams
  - https://www.kdnuggets.com/2016/10/battle-data-science-venn-diagrams.html

Section #3 - Impacts, Importance and examples

Lecture #19 - Why is this important now?

- Fortune / Facebook Is Now More Valuable Than Exxon / by Claire Zillman / Feb 1, 2016
  - https://fortune.com/2016/02/01/facebook-value-exxon/

- Note to Self / Digging Into Facebook's File on You / Sept 27, 2016
  - https://www.wnycstudios.org/podcasts/notetoself/episodes/propublica-facebook-algorithms-bias-privacy

- Moore's law Definition from Wikipedia, the free encyclopedia
  - https://en.wikipedia.org/wiki/Moore%27s_law

- YCOMBINATOR 2014 DATA SCIENCE START-UPS
  - https://mlwave.com/ycombinator-2014-data-science-start-ups/

- Data Scientist: The Sexiest Job of the 21st Century / Harvard Business Review / Oct 2012
  - https://hbr.org/2012/10/data-scientist-the-sexiest-job-of-the-21st-century

- McKinsey Report Highlights the Impending Data Scientist Shortage / Paul M. Davis
  - https://tanzu.vmware.com/content/blog/mckinsey-report-highlights-the-impending-data-scientist-shortage

Lecture #20 - Computers exploding! - The explosive growth of computer power explained.

- Processing Power Compared *(link updated)*
  - https://insightaas.com/infographic-processing-power-compared-1956-2015-experts-exchange/

- Supercomputer Timeline
  - https://mason.gmu.edu/~tbell5/page2.html

- FLOPS Definition from Wikipedia, the free encyclopedia (link updated)
  - https://en.wikipedia.org/wiki/FLOPS

- Prices in the 80's
  - http://www.inthe80s.com/prices.shtml

- USA Today/ Average new car price zips 2.6% to $33,560 / May 4, 2015
  - https://www.usatoday.com/story/money/cars/2015/05/04/new-car-transaction-price-3-kbb-kelley-blue-book/26690191/

- US Inflation Calculator
  - https://www.usinflationcalculator.com/

Lecture #21 - What problems does Machine Learning Solve?

- Regression Analysis Tutorial and Examples
  - https://blog.minitab.com/blog/adventures-in-statistics-2/regression-analysis-tutorial-and-examples

Lecture #22 - Where it's transforming our lives

- YCOMBINATOR 2014 DATA SCIENCE START-UPS
  - https://mlwave.com/ycombinator-2014-data-science-start-ups/

- Radio Lab / Million Dollar Microsecond / Feb 4, 2013
  - https://www.wnycstudios.org/podcasts/radiolab/segments/267195-million-dollar-microsecond

- Mining for Causal Relationships: A Data-Driven Study of the Islamic State
  - https://arxiv.org/pdf/1508.01192v1.pdf

- Campaign to Stop Killer Robots
  - https://www.stopkillerrobots.org/

Section #4 - The Machine Learning Process

Lecture #26 - 2 - Identifying, obtaining, and preparing the right data

- Datasets for Data Mining, Data Science, and Machine Learning
  - https://www.kdnuggets.com/datasets/index.html

- Largest API Directory on the Web
  - https://www.programmableweb.com/apis/directory

- Data Science Toolkit
  - http://www.datasciencetoolkit.org/

- Quora - Which are some of the best web data scraping tools?
  - https://www.quora.com/What-are-some-of-the-best-web-data-scraping-tools

- 5 Web Scraping Tools to Extract Online Data
  - http://www.hongkiat.com/blog/web-scraping-tools/

- Tidy Data by Hadley Wickham
  - http://vita.had.co.nz/papers/tidy-data.pdf

Section #5 - How to apply Machine Learning for Data Science

Lecture #32 - Common platforms and tools for Data Science

- Computing Platforms for Analytics, Data Mining, Data Science
  - https://www.kdnuggets.com/2015/04/computing-platforms-analytics-data-mining-data-science.html

- R, Python Duel As Top Analytics, Data Science software – KDnuggets 2016 Software Poll Results
  - https://www.kdnuggets.com/2016/06/r-python-top-analytics-data-mining-data-science-software.html

Lecture #33 - Data Science using – R

- The R Project for Statistical Computing
  - https://www.r-project.org/

- The "CRAN" Comprehensive R Archive Network
  - https://cran.r-project.org/

- Rstudio
  - https://rstudio.com/

- Jupyter Notebook
  - https://jupyter.org/

Lecture #34 - Data Science using – Python

- Python.org
  - https://www.python.org/

- Jupyter Notebook
  - https://jupyter.org/

- Python Idle
  - https://docs.python.org/3/library/idle.html

- Top 20 Python Machine Learning Open Source Projects
  - https://www.kdnuggets.com/2015/06/top-20-python-machine-learning-open-source-projects.html

- CONDA (lets you switch easily between PY versions)
  - https://docs.conda.io/en/latest/

- Continuum Anaconda (free) Distribution
  - https://www.anaconda.com/products/individual

- Enthought Canopy (free and pro) Distribution [Now discontinued]
  - https://assets.enthought.com/downloads/

Lecture #35 - Data Science using SQL

- SQL Definition from Wikipedia, the free encyclopedia
  - https://en.wikipedia.org/wiki/SQL

- MySQL (Open source implementation of SQL)
  - https://www.mysql.com/

- PostgreSQL (Open source implementation of SQL)
  - https://www.postgresql.org

- Microsoft SQL Server (Paid implementation of SQL)
  - https://www.microsoft.com/en-us/sql-server

- Oracle Database (Paid implementation of SQL)
  - https://www.oracle.com/database/

- phpMyAdmin (Open source tool used to help compose the SQL language into queries)
  - https://www.phpmyadmin.net/

- Toad for SQL server (Paid tool used to help compose the SQL language into queries)
  - https://www.quest.com/products/toad-for-sql-server/

Lecture #36 - Data Science using Excel

- Microsoft Azure Machine Learning Excel Add-in
  - https://azuremlexcel.codeplex.com/

- Open Office Calc (Open source tool similar to excel)
  - https://www.openoffice.org/product/calc.html

Lecture #37 - Data Science using RapidMiner

- RapidMiner's data science platform
  - https://rapidminer.com/