

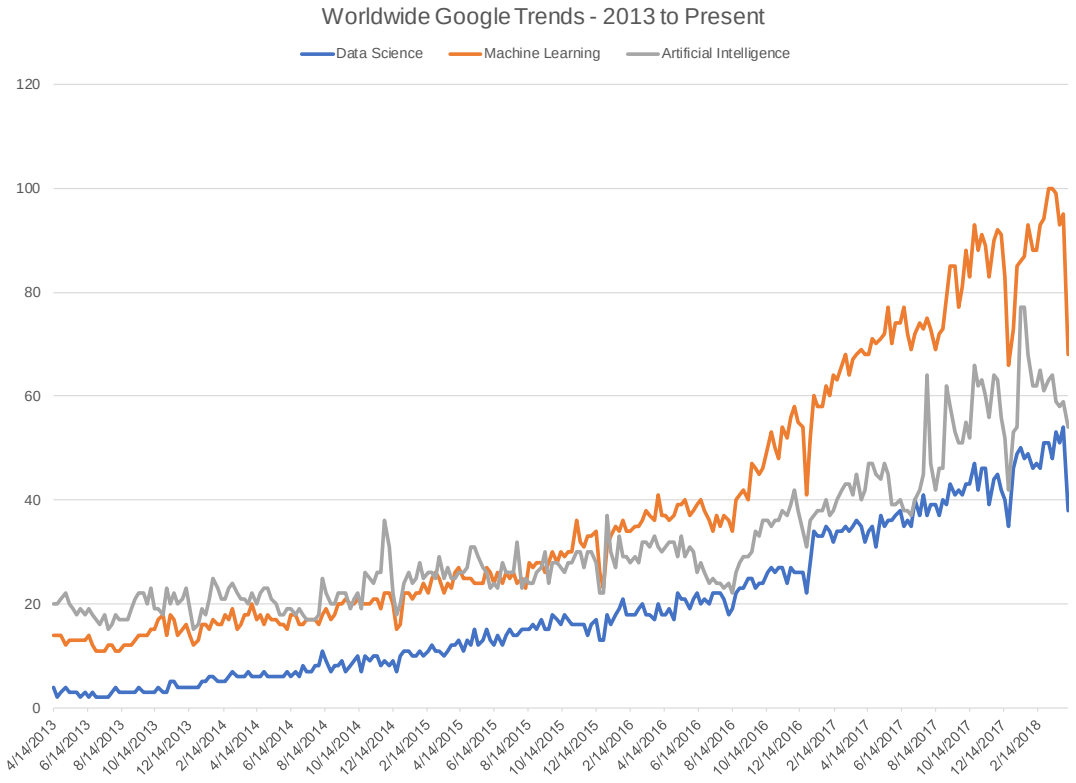
R Workshop – Sarah Gets a Diamond

April 11, 2018

Sponsors: Data Science Club & Technology Club
Presenter: Dang Trinh

Table of Contents

Data science? Machine learning? Artificial intelligence?



SPOTLIGHT ON BIG DATA

Data Scientist: The Sexiest Job Of the 21st Century

Meet the people who can coax treasure out of messy, unstructured data.
by Thomas H. Davenport and D.J. Patil

What AI can and can't do (yet) for your business

Artificial intelligence is a moving target. Here's how to take better aim.

by Michael Chui, James Manyika, and Mehdi Miremadi

The data science Venn diagram – Drew Conway

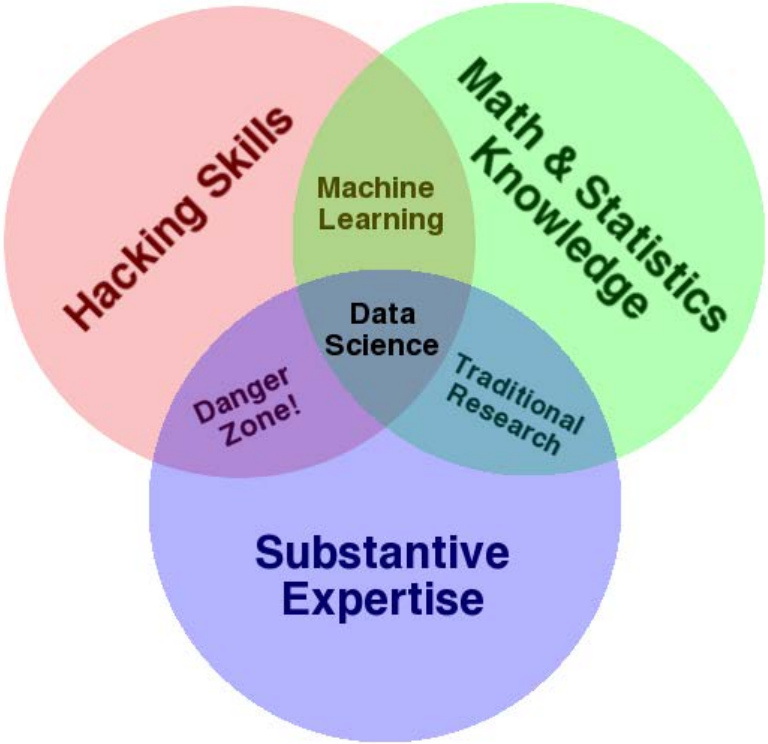


Table of Contents

Reproducibility & self-containment

Table of Contents

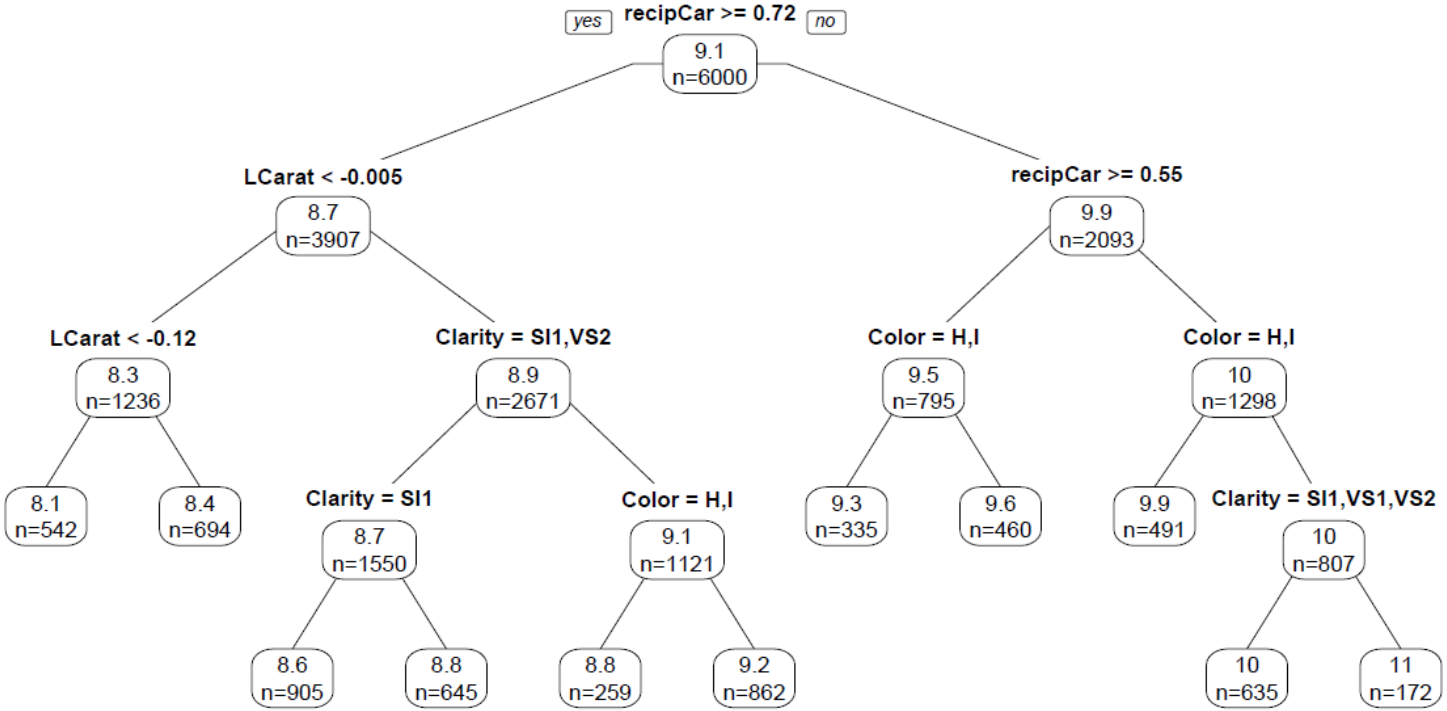
Table of Contents

Table of Contents

Table of Contents

Table of Contents

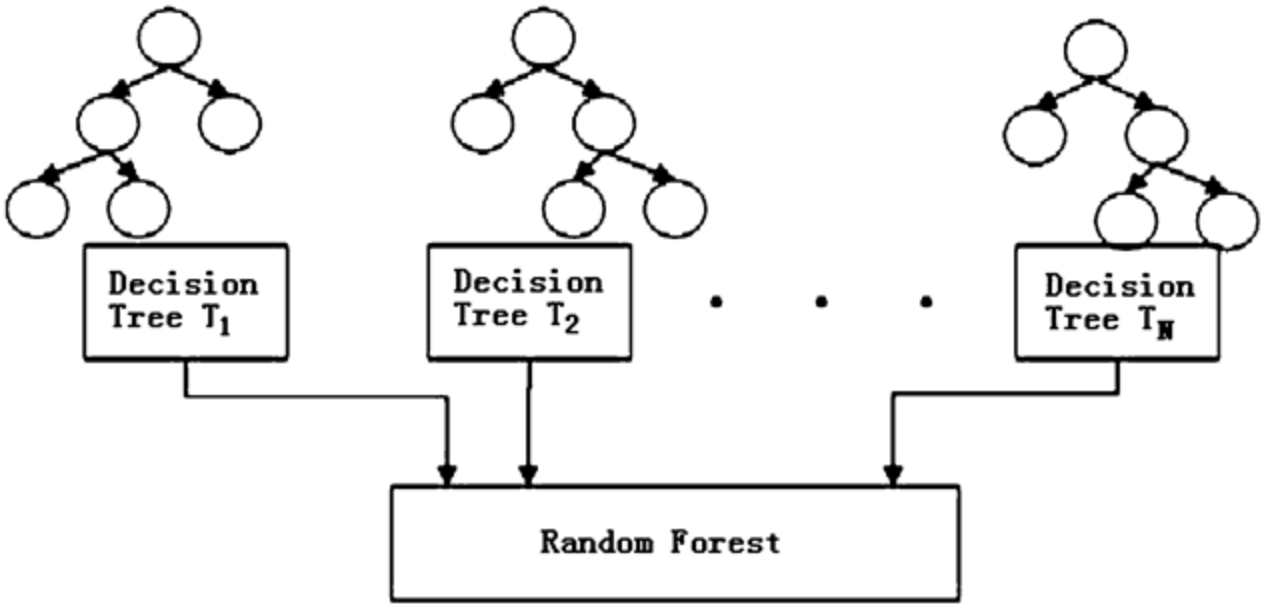
Regression trees/classification trees



Ensembles

	Correlation	
Accuracy	High	Low
High		
Low		

Bagged Trees & Random Forest



Boosted Trees

LASSO (least absolute shrinkage and selection operator)

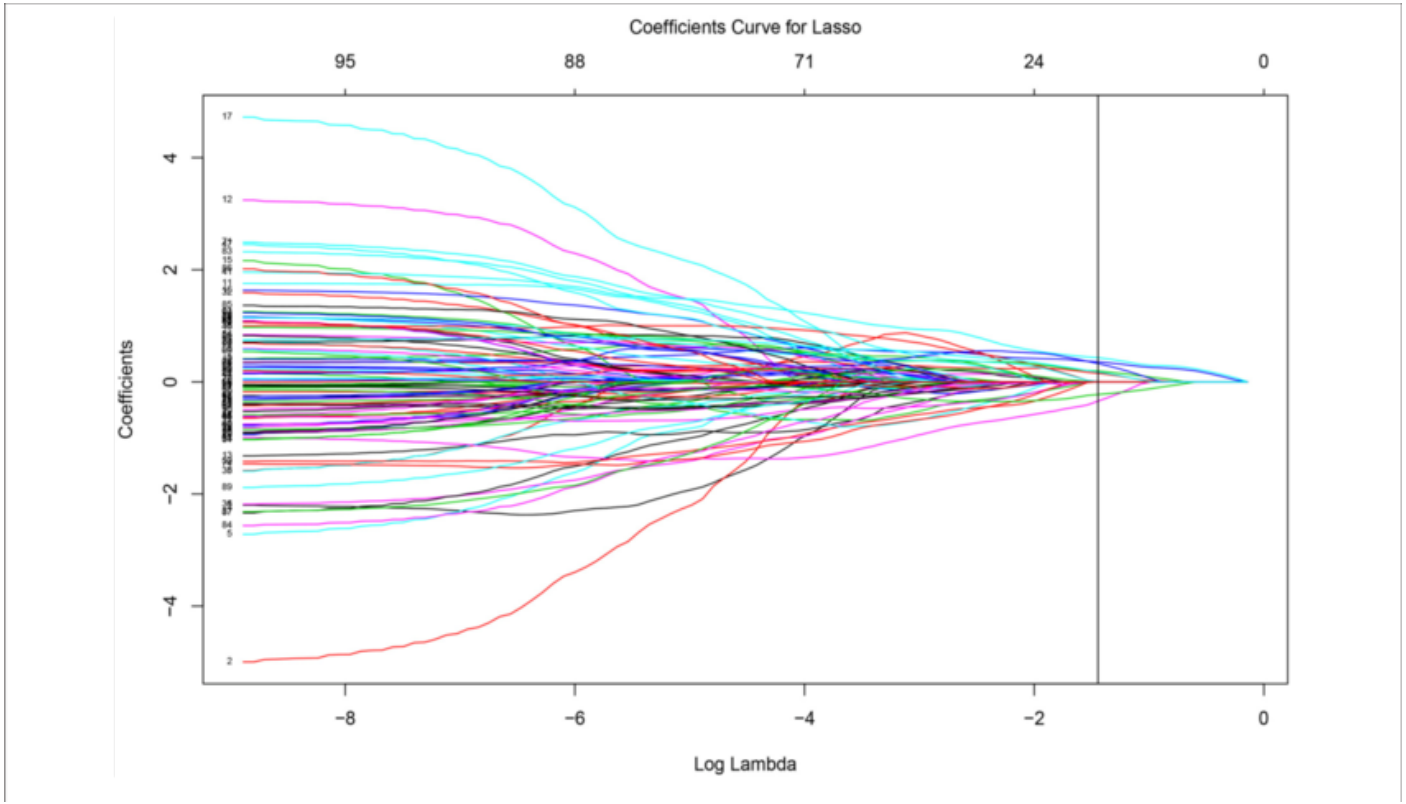


Table of Contents

Highly recommended – extensive materials

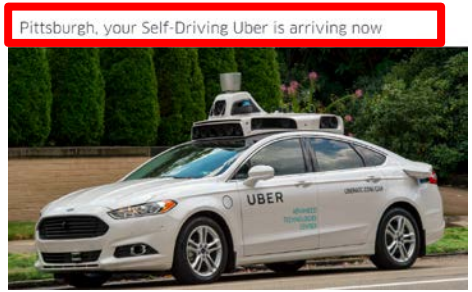
GBUS 8496: Data Science in Business, 2017-2018 Quarters 1 and 2, Course Outline

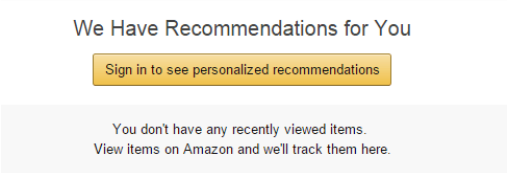
Module	Class	Date	Case/Deliverable	Topic
Introduction	1	Aug 23-W	Carvana I	Visual Analytics, Tableau
Time Series	2	Aug 24-Th	NICU (A)	Additive Trends
Time Series	3	Aug 30-W	Tumblr	Multiplicative Trends
Time Series	4	Aug 31-Th	NICU (B)	Seasonality and Distributional Trends
Time Series	5*	Sep 6-W	Split Banana	Multiple Seasonalities
Time Series	6	Sep 7-Th	Sabermetrics & SPF	Time Series & Wisdom of Crowds
Advanced Regression	7	Sep 13-W	Casino Jack	Interactions, Optimization
Advanced Regression	8	Sep 14-Th	Clinton-Obama I	Regression Trees, Regularization
Advanced Regression	9*	Sep 20-W	Clinton-Obama II	More Regression Trees, Regularization
Machine Learning	10	Sep 21-Th	Bike Sharing	Random Forests
Data Wrangling	11	Sep 27-W	FanShop	Basic Joins and Filters
Data Wrangling	12	Sep 28-Th	Fannie Mae I	Advanced Joins
Machine Learning	13	Oct 4-W	Fannie Mae II	Boosted Trees
Machine Learning	14*	Oct 5-Th	Fannie Mae III	Comprehensive
Integration	-	Oct 10-Tu	Individual Project	Comprehensive
Machine Learning	15*	Oct 18-W	Carvana II	Stacking
Machine Learning	16	Oct 19-Th	MovieLens I	Recommender Systems, Spark
Machine Learning	17	Oct 25-W	MovieLens II	Segmentation, Design Thinking
Machine Learning	18	Oct 26-Th	Digit Recognizer	Neural Networks, Image Recognition
Machine Learning	19	Nov 1-W	Dogs vs. Cats	Image Recognition, Tensorflow
Machine Learning	20	Nov 2-Th	Gender Recognition by Voice	Voice Recognition
NLP	21	Nov 8-W	Twitter I	Document-Term Matrix, Word Clouds
NLP	22	Nov 9-Th	Twitter II	Sentiment Analysis, Network Analysis
NLP	23	Nov 15-W	Being Presidential	Sentiment Analysis, Category Dictionaries
NLP	24*	Nov 16-Th	Being Presidential	Sentiment Analysis, Category Dictionaries
Integration	-	Nov 20-M	Group Project Proposal	Comprehensive
NLP	25	Nov 29-W	Yelp	Semantic Analysis, Topic Modeling
NLP	26	Nov 30-Th	RateBeer	Semantic Analysis, Topic Modeling
Integration	27	Dec 6-W	Group Project Presentations	Comprehensive
Integration	28	Dec 7-Th	Group Project Presentations	Comprehensive

Table of Contents

AI Undoubtedly Revolutionized Our Lives & Remains at the Forefront of Our Future

- Though most people do not realize it, our daily lives are inundated with products made possible only through artificial intelligence





Yet Subconscious Bias Managed to Bury Its Claws into this Future's Frontier

- "Histories of discrimination can live on in digital platforms, and if they go unquestioned, they become part of the logic of everyday algorithmic systems"

DIGITS

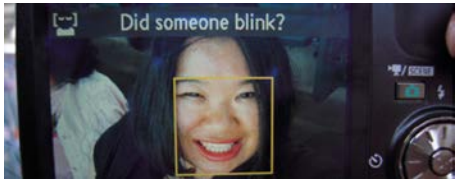
Google Mistakenly Tags Black People as 'Gorillas,' Showing Limits of Algorithms

Tuesday, July 7, 2015

QUESTIONING THE FAIRNESS OF TARGETING ADS ONLINE

CMU Probes Online Ad Ecosystem

Machine Bias: There's Software Used Across the Country to Predict Future Criminals. And it's Biased Against Blacks. – ProPublica



With Great Power Comes Great Responsibility

Kate Crawford, "AI's White Guy Problem," *NY Times*

- Just like any tool created by human, AI suffers from the faults of its creator
- Therefore, as creator of AI, we must be wary of and effectively address our own shortcomings

*"We need to be vigilant about how we design and train these machine learning systems, or we will see **ingrained forms of bias** built into the artificial intelligence of the future"*

*"If we look at how systems can be discriminatory now, we will be much better placed to design fairer artificial intelligence. But that requires **far more accountability** from the tech community"*

*"While machine learning technology can offer unexpected insights and new forms of convenience, we must address the current implications for **communities that have less power**, for those who aren't dominant in elite Silicon Valley circles"*