



the Machine Learning starter pack

starting and moving forward with ML

Jordan Aiko Deja

Center for Complexity and Emerging Technologies (ComET)

Philippines Computer Human Interaction Group (PhiCHI)



COMET

TWEET TO ME @jordandoinwork EMAIL ME jordan_deja@dsu.edu.ph VISIT ME <http://spark.adobe.com/page/A1gpP/>



hi, I'm Jordan!



COMET

i'm part of a lab where
we do research in **HCI**
using ML techniques

TWEET TO ME @jordanoinwork

EMAIL ME jordan_deja@du.edu.ph

VISIT ME <http://spark.adobe.com/page/A1gpP/>

machine learning.



“

machine learning is a subfield of computer science. it gives computers the ability to learn without being explicitly programmed.”

-Arthur Samuel



you can program a computer to do a certain task. if there's a special condition, we program a **then** for that **if**. the problem is **the “what-ifs”** that we encounter. do we flood our programs with what if's? to teach a computer context, do we have to feed it with **billions of rules?**

at least that's what Cycorp did before. You can now check OpenCyc.

brute force is just bad.

TWEET TO ME @jordandejaj

EMAIL ME jordan.deja@du.edu.ph

VISIT ME <http://spark.adobe.com/page/A1gpP/>

6 ML startup tips.



COMET

tip#01

machine learning is simply **not** a new name for
artificial intelligence

TWEET TO ME @jordandoinwork EMAIL ME jordan.deja@disu.edu.ph VISIT ME <http://spark.adobe.com/page/A1gpP/>

machine learning

Web Images News More ▾ Search tools

About 138,000,000 results (0.29 seconds)

Machine learning - Wikipedia, the free encyclopedia
en.wikipedia.org/wiki/Machine_learning ▾
Machine learning, a branch of artificial intelligence, concerns the construction and study of algorithms and systems that can learn from data. In general, Machine ...
Artificial intelligence - Supervised learning - List of machine learning ... - Weka

News for machine learning
Why machine learning might be a wearable camera's best friend
GigaOM - 4 days ago
The glut of research in teaching computers to analyze and understand images could prove very helpful in letting us take full advantage of the ...
Machine Learning / Scientific Software Engineer - Plymouth
This is Plymouth - 12 hours ago

Machine Learning | Coursera
<https://www.coursera.org/course/ml> ▾
Machine learning is the science of getting computers to act without being explicitly programmed. In the past decade, machine learning has given us self-driving ...
Edward Choi +1'd this

Machine Learning - Springer
www.springer.com/home/computer-science/artificial-intelligence ▾
Machine Learning is an international forum for research on computational approaches to learning. The journal publishes articles reporting substantive results on ...

Machine Learning (Theory)



“ A computer program is said to learn from an experience E with respect to some task T and some performance measure P , If its performance on T as measured by P , improves by some experience E.

- Tom Mitchell



TWEET TO ME @jordanndoingwork

EMAIL ME jordan.deja@disu.edu.ph

VISIT ME <http://spark.adobe.com/page/A1gpP/>

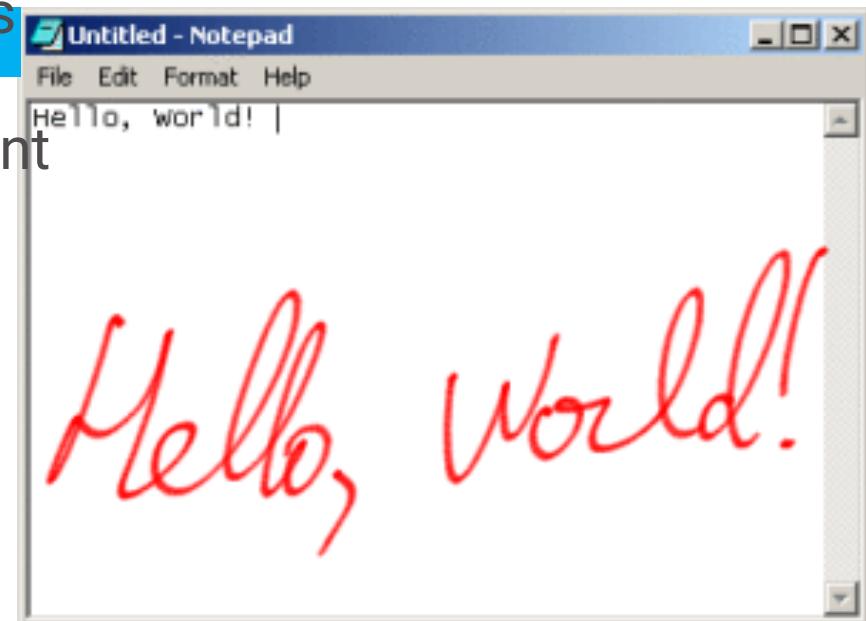
Learning to Play Checkers

- **Task:** playing checkers
- **Performance measure:** percent of games won against opponents
- **Experience:** playing practice games against itself



Handwriting Recognition

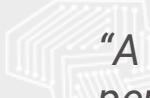
- **Task:** recognizing and classifying handwritten words within images
- **Performance measure:** percent of words correctly classified
- **Experience:** database of handwritten words with given classifications



Robot Driving

- **Task:** driving on public four-lane highways using vision sensors
- **Performance measure:** average distance travelled before an error is encountered (judged by a human overseer)
- **Experience:** sequences of images and steering commands recorded while observing a human driver





COMET

"A computer program is said to learn from experience E with respect to some task T and some performance measure P , if its performance on T , as measured by P , improves with experience E ."

Suppose your email program watches which emails you do or do not mark as spam, and based on that learns how to better filter spam. What is the task T in this setting?

- T** A. Classifying emails as spam or not spam.
- E** B. Watching you label emails as spam or not spam.
- P** C. The number (or fraction) of emails correctly classified as spam/not spam.
- D. None of the above—this is not a machine learning problem.



tip#02

know that machine learning and data mining
are not entirely the same.



COMET

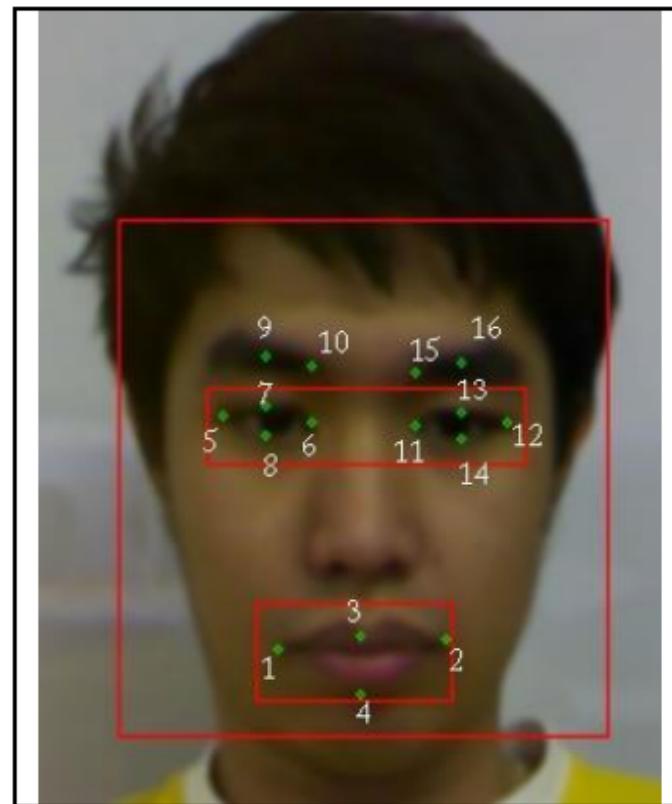
“machine learning focuses on prediction on **known** properties learned from training data. however, data mining focuses on the discovery of **previously-unknown** properties in the data.

TWEET TO ME @jordandoinwork

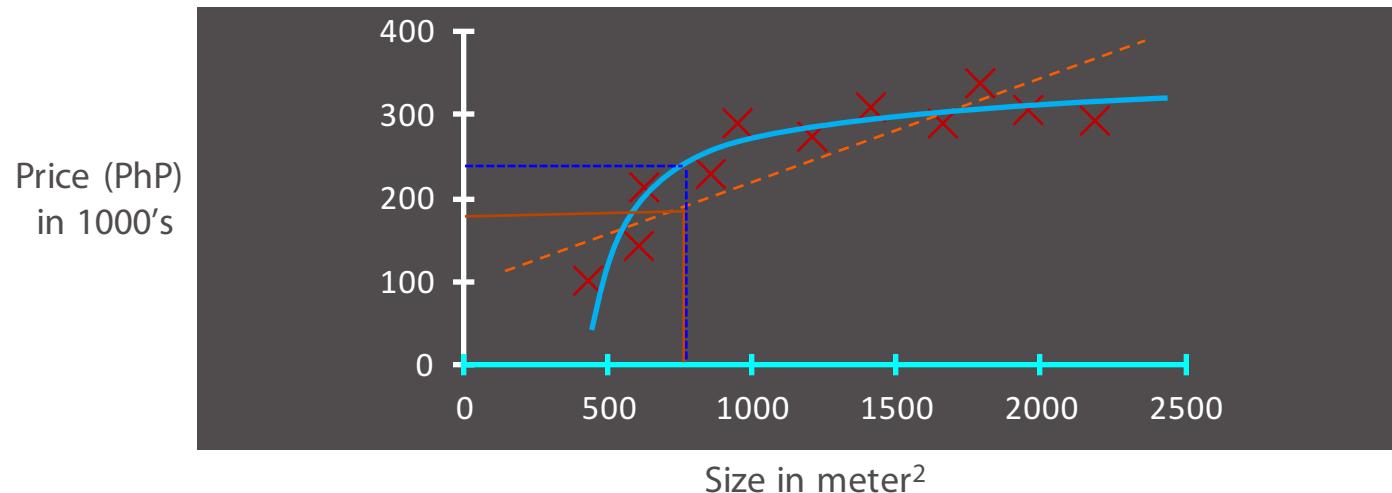
EMAIL ME jordan.deja@disu.edu.ph

VISIT ME <http://spark.adobe.com/page/A1gpP/>

how do we teach the computer how a sad face looks like?



Housing price prediction

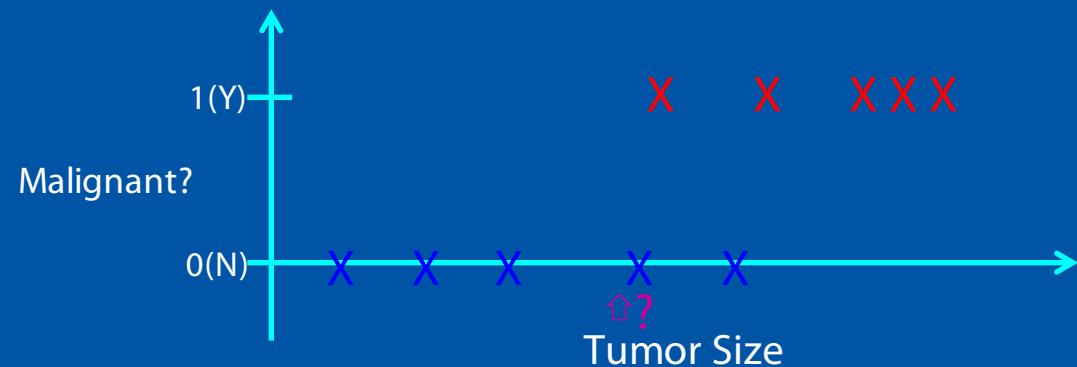


Supervised Learning

“right answers” given for each example in the data

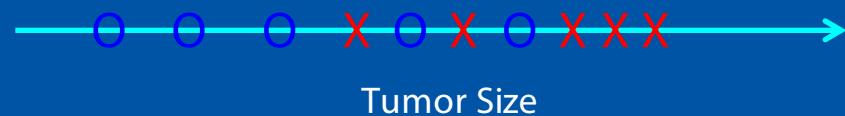
Regression: Predict continuous valued output (price)

Breast cancer (malignant, benign)



Classification
Predict a
discrete valued
output (0 or 1)

Another representation (using 1 feature):





COMET

You're running a company, and you want to develop learning algorithms to address each of two problems.

Problem 1: You have a large inventory of identical items. You want to predict how many of these items will sell over the next 3 months.

Problem 2: You'd like software to examine individual customer accounts, and for each account decide if it has been hacked/compromised.

Should you treat these as classification or as regression problems?

- A. Treat both as classification problems.
- B. Treat problem 1 as a classification problem,
problem 2 as a regression problem.
- C. Treat problem 1 as a regression problem,
problem 2 as a classification problem.
- D. Treat both as regression problems.



COMET

TWEET TO ME @jordandoinwork

EMAIL ME jordan_deja@dsu.edu.ph

VISIT ME <http://spark.adobe.com/page/A1gpP/>

“learning is experience.
Everything else is just
information” — Einstein



photo from google images



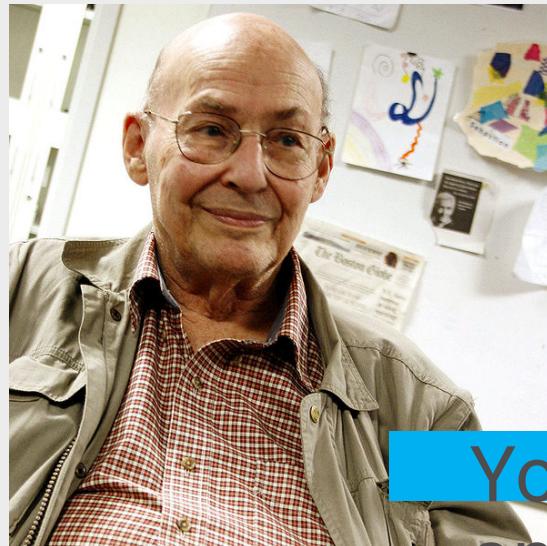
COMET

tip#03

machine learning requires a lot of data



COMET



You don't understand
anything until you learn
it more than one way
- Marvin Minsky



COMET

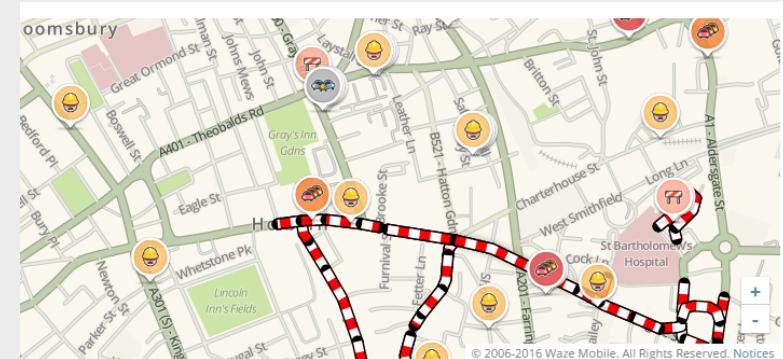
we collected user data of 40 students. they used a tool for 10 hours each. one second consisted one record in the data set. that's around 1.4M rows. we were able to generate a multi-layer perceptron with a 59% accuracy.

we had around 22 features per row

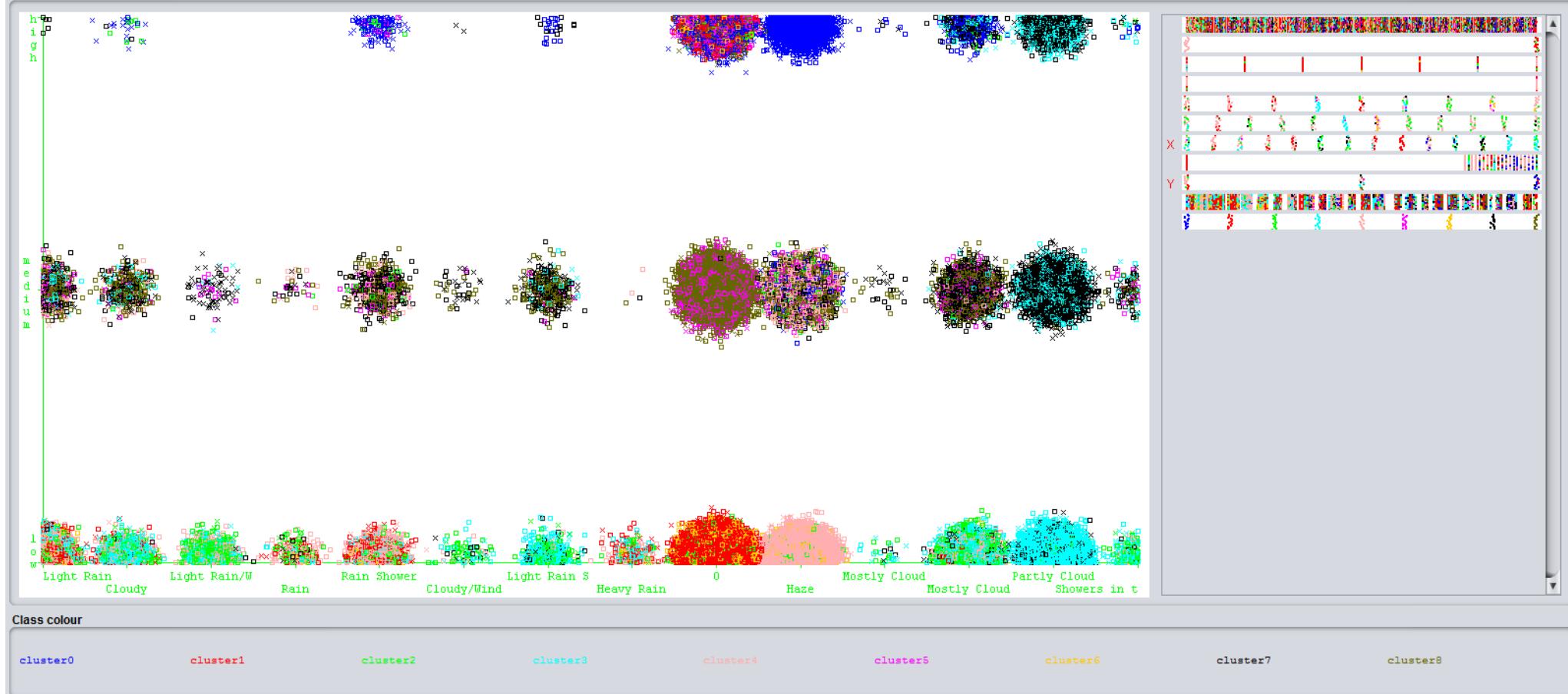


COMET

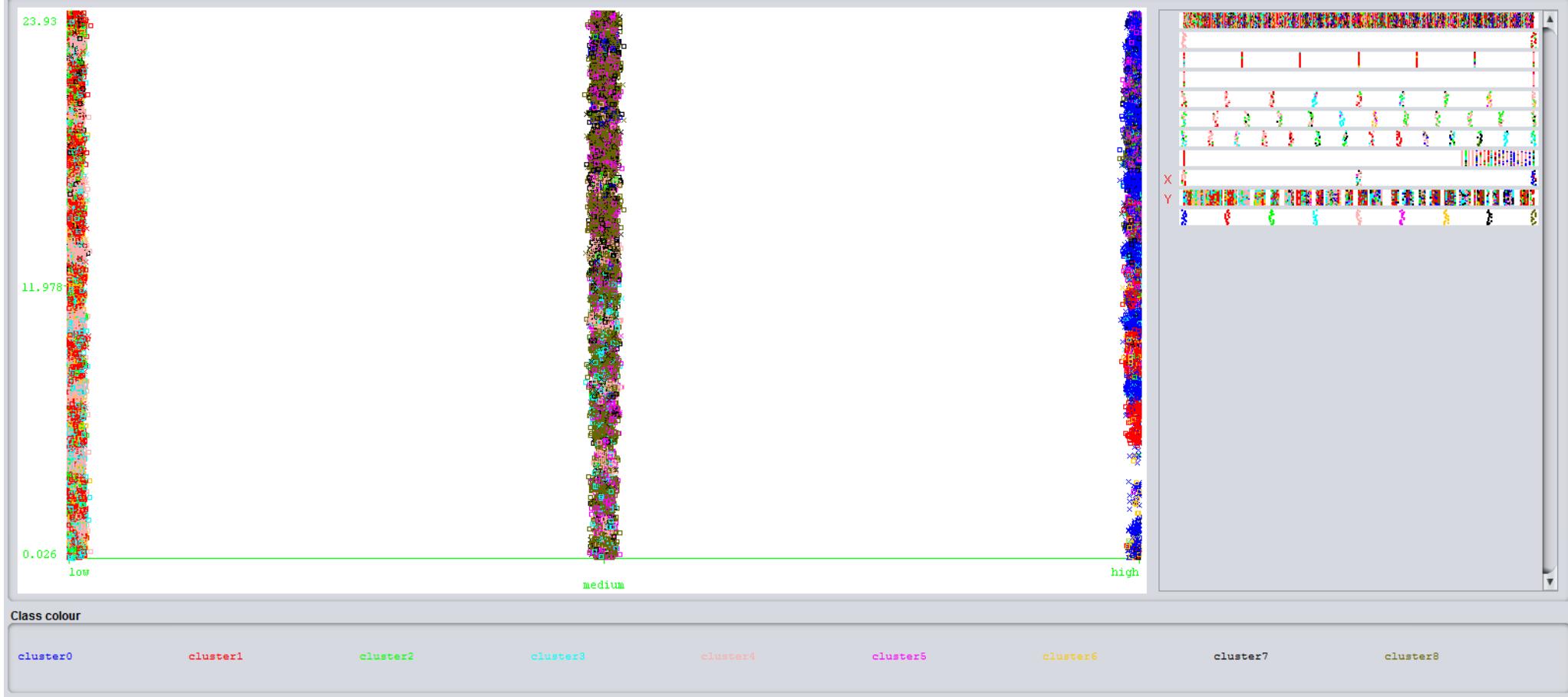
we worked with 25k rows worth of data. we were trying to find congestion patterns in certain traffic points in MMDA in all its cities. and this was just for the first 2 months.



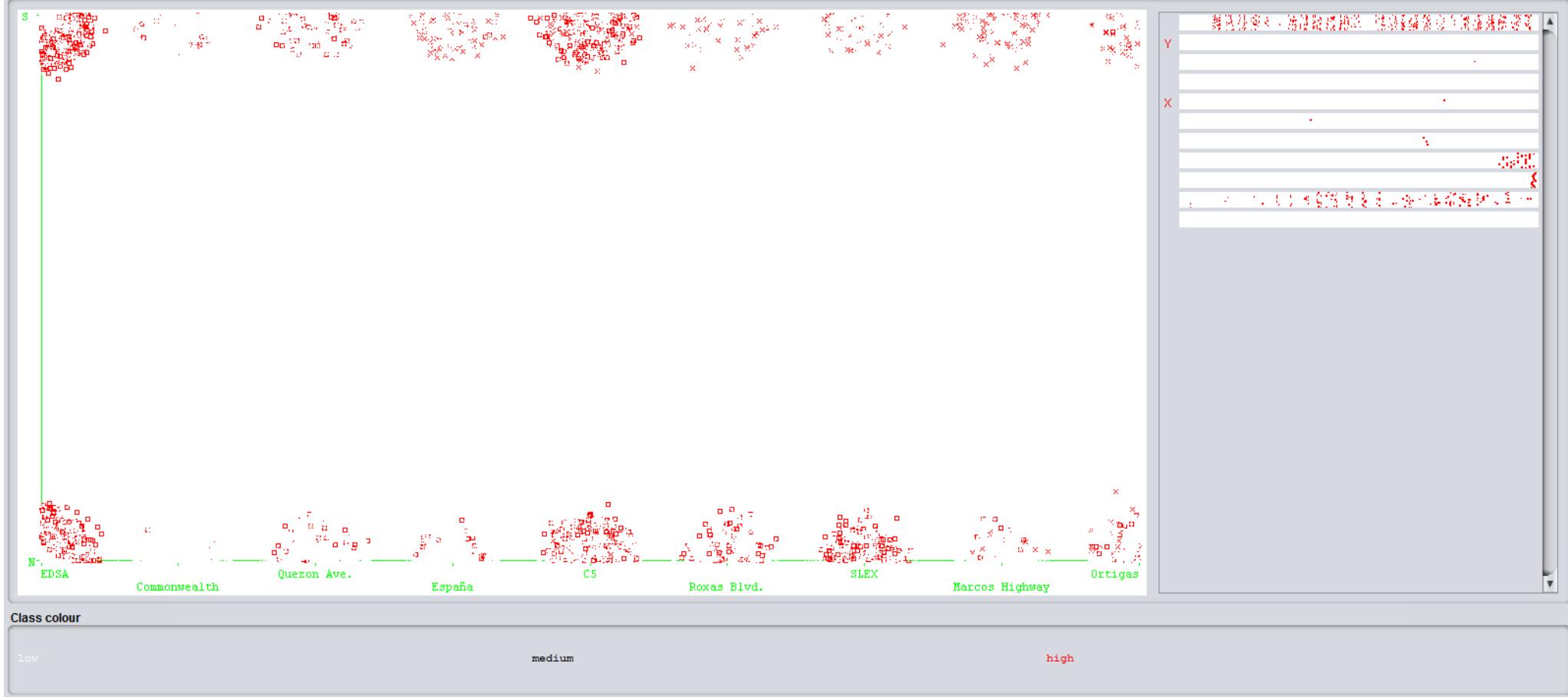
Plot: MASTER-weka.filters.unsupervised.attribute.Remove-R1,4_clustered



Plot: MASTER-weka.filters.unsupervised.attribute.Remove-R1,4_clustered



Plot: MASTER-weka.filters.unsupervised.attribute.Remove-R1,4_clustered



Class colour

low medium high

curse of dimensionality

“as the data becomes more sparse it becomes more difficult to find significance in our data”



COMET

tip#04

in ML you are at **the mercy** of your data

TWEET TO ME @jordandoinwork

EMAIL: ME jordan.deja@duis.edu.ph

VISIT ME <http://spark.adobe.com/page/A1cpP/>

Good data means good results



COMET

to ensure good data, it's best to gather data following well-established protocols and procedures. we don't want data to be overfit or to be pulled off from chance aka "tsamba".

we also need a good understanding of statistics to be able to find relationships in these numbers.



COMET

tip#05

machine learning **is not a thing** of the future. It's here



TWEET TO ME @jordandoinwork

EMAIL ME jordan_deja@dsu.edu.ph

VISIT ME <http://spark.adobe.com/page/A1gpP/>

“

thinking is a human feature. Will
AI someday really think? That's like
asking if submarines swim. If you
call it swimming, then robots will
think, yes!"

-Noam Chomsky



we already have Watson

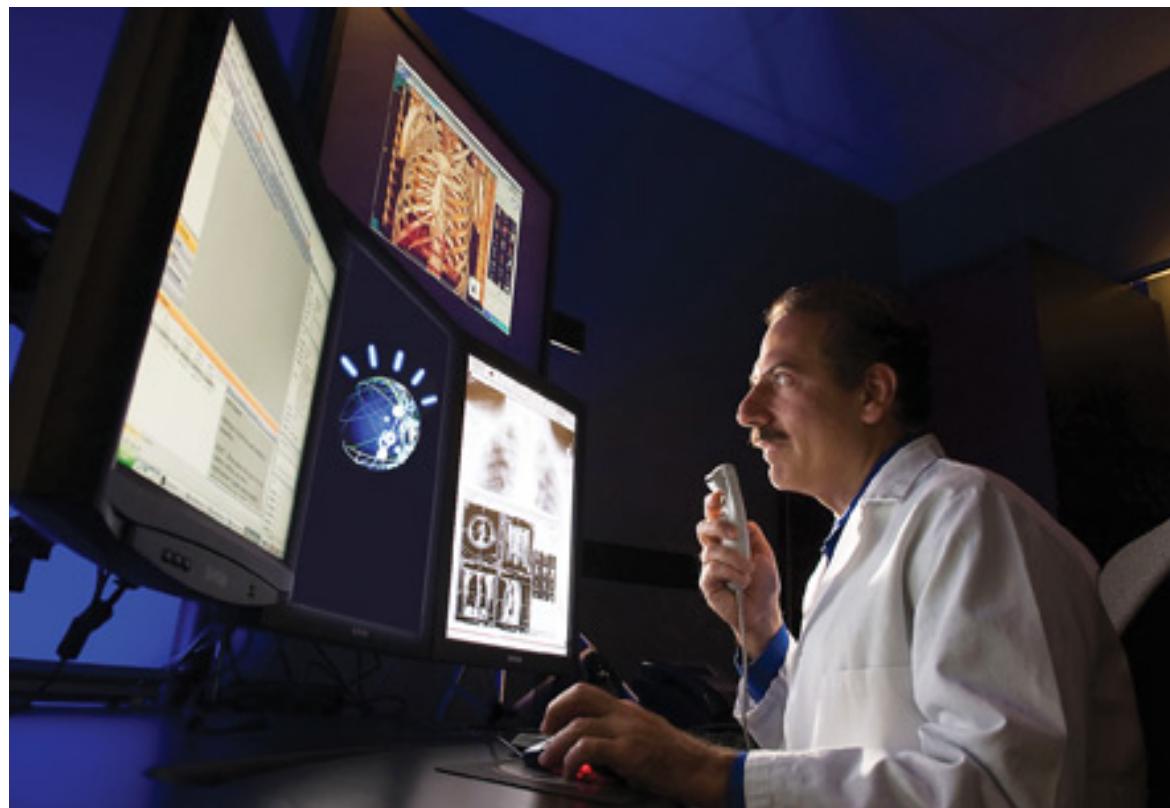


TWEET TO ME @jordandoinwork

EMAIL ME jordan.deja@disu.edu.ph

VISIT ME <http://spark.adobe.com/page/A1gpP/>

Watson is already used in healthcare



TWEET TO ME @jordandoinwork

EMAIL ME jordan.deja@disu.edu.ph

VISIT ME <http://spark.adobe.com/page/A1gpP/>

Ada and Grace as virtual museum agents



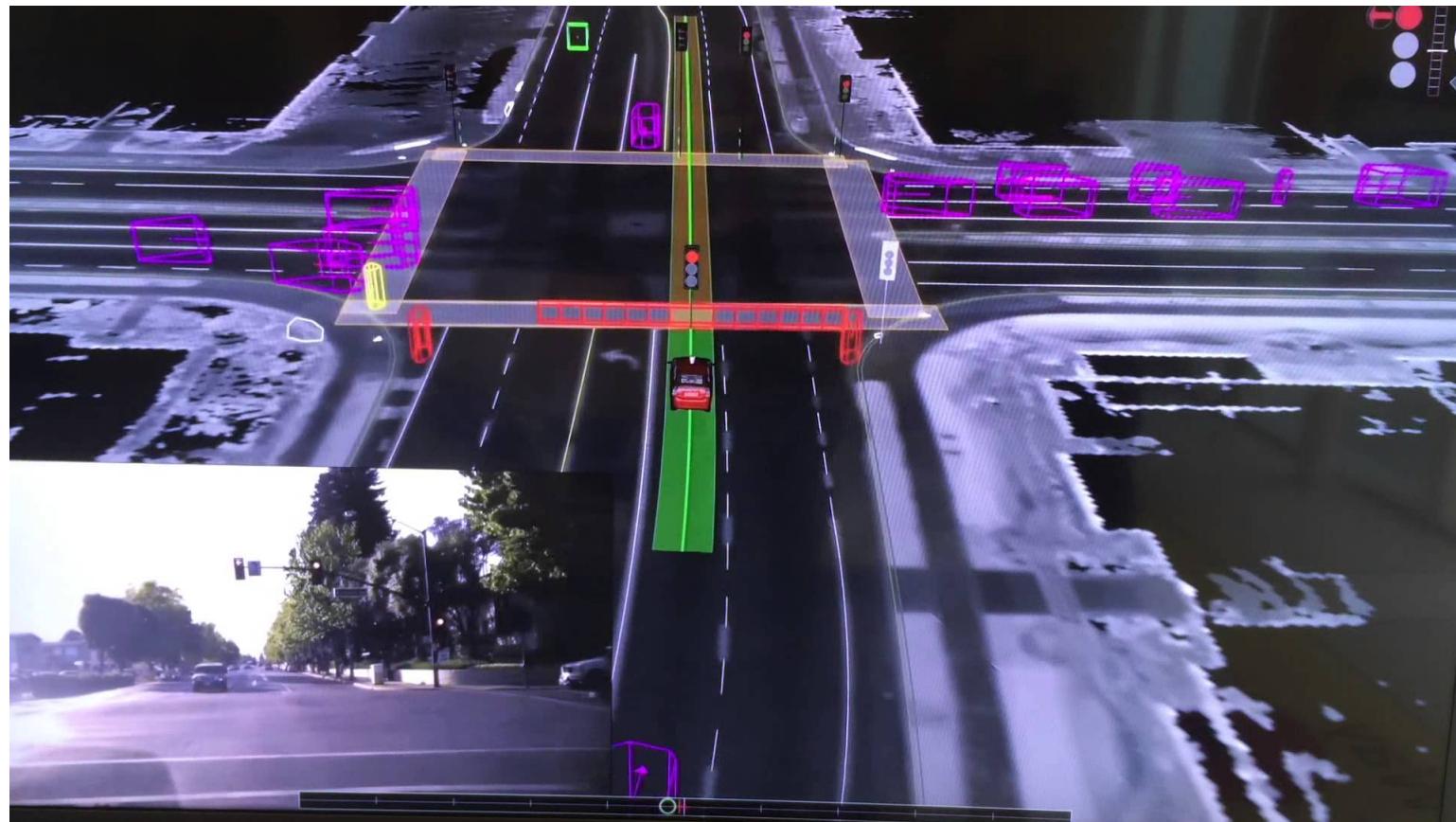
photo from virtualagentchat.com

TWEET TO ME @jordandoinwork

EMAIL ME jordan.deja@disu.edu.ph

VISIT ME <http://spark.adobe.com/page/A1gpP/>

Google's self driving car



GOOGLE I/O 2016 EXTENDED MANILA

TWEET TO ME @jordandoinwork

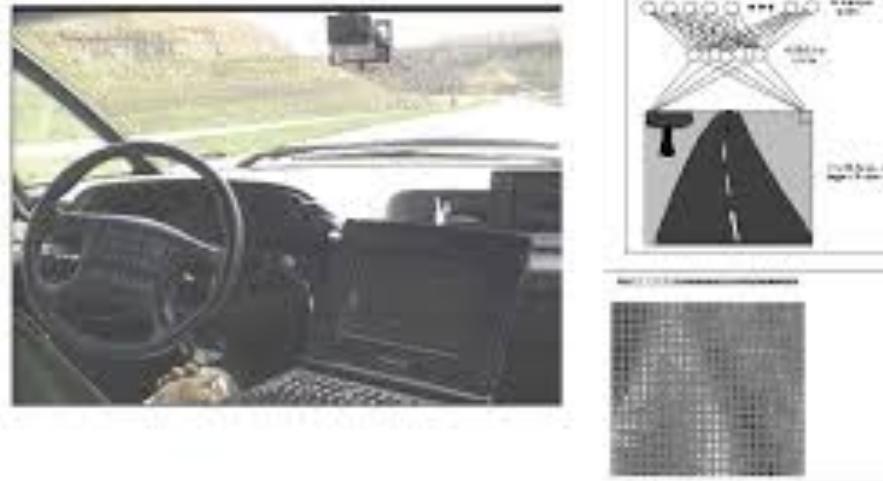
EMAIL ME jordan.deja@disu.edu.ph

VISIT ME <http://spark.adobe.com/page/A1gpP/>

ALVINN – the first vehicle to cross from West to East Coast of the US, unmanned (1994)

Neural Net example: ALVINN

- Autonomous vehicle controlled by Artificial Neural Network
- Drives up to 70mph on public highways



Note: most images are from the online slides for Tom Mitchell's book "Machine Learning".

TWEET TO ME @jordandoinwork

EMAIL ME jordan.deja@disu.edu.ph

VISIT ME <http://spark.adobe.com/page/A1gpP/>

reading of muscle signals (electromyography) to recreate common gestures

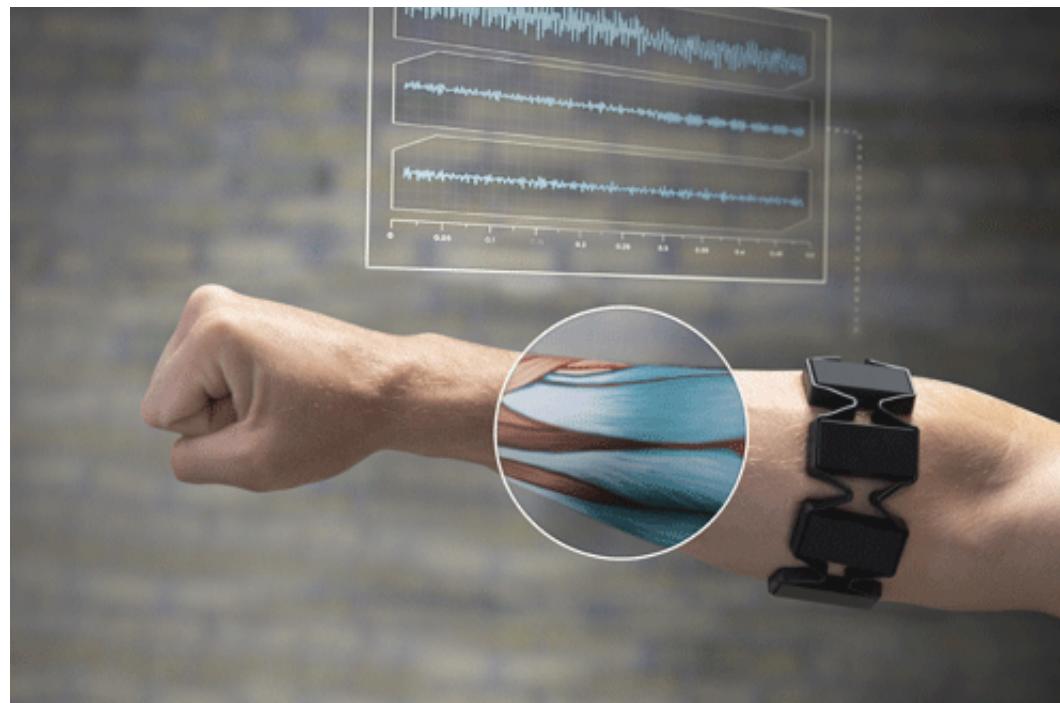


photo from thalmic labs



tip#06

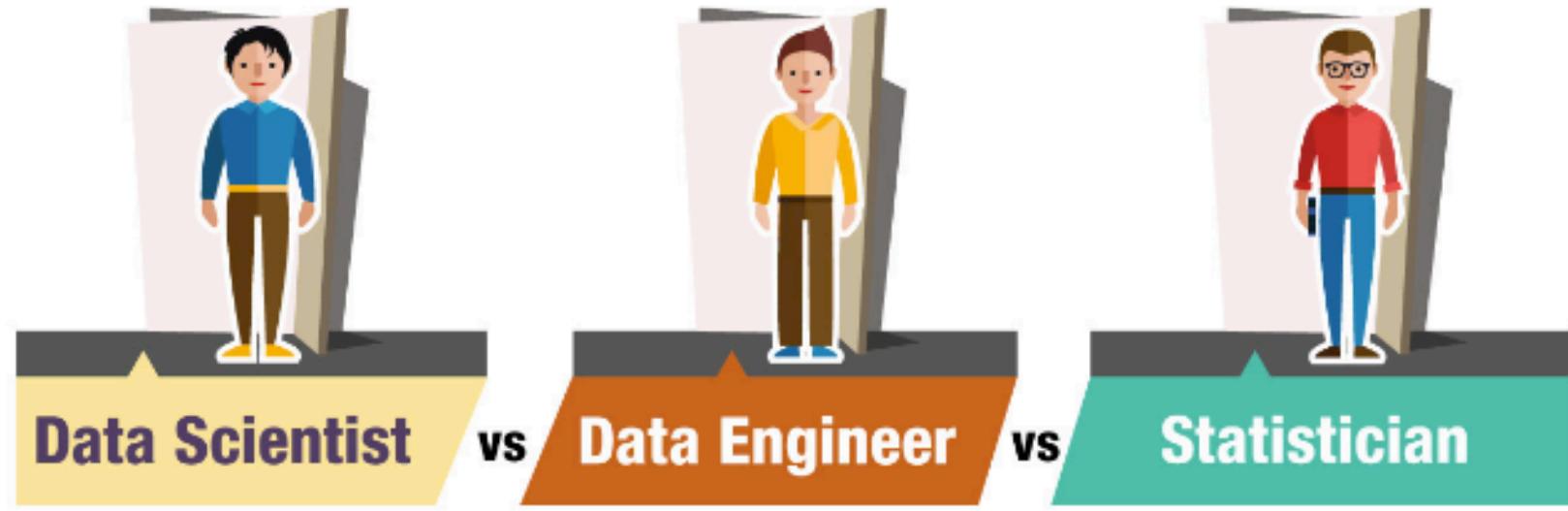
okay machine is learning is there. there are jobs
already waiting for it

Yes there are ML-jobs even in the PH

ML is a jargon. ML-jobs hide under names like Data Scientist, Analytics Engineer. You need experience using ML tools like WEKA, RapidMiner, etc.

TWEET TO ME @jordandoinwork

EMAIL ME jordan.deja@disu.edu.ph

VISIT ME <http://spark.adobe.com/page/A1gpP/>

TWEET TO ME @jordandoinwork

EMAIL ME jordan.deja@disu.edu.ph

VISIT ME <http://spark.adobe.com/page/A1gpP/>

Data Scientist vs Data Engineer vs Statistician

Responsibilities

- Develop and plan required analytic projects in response to business needs.
- Contribute to data mining architectures, modeling standards, reporting, and data analysis methodologies.
- Collaborate with stakeholders to integrate data mining results with existing systems.
- Monitor data mining system performance and implement efficiency improvements.

Skills

- Programming, Mathematics, Business Understanding, Statistics, Data Visualization, Machine Learning, Attention to detail

- Design, construct, install, test and maintain highly scalable data management systems
- Improve data foundational procedures, guidelines and standards
- Integrate new data management technologies and software engineering tools into existing structures
- Create custom software components (e.g. specialized UDFs) and analytics applications

- Apply statistical theories and methods to solve practical problems of various industries
- Determine methods for finding or collecting data
- Design surveys or experiments or opinion polls to collect data
- Analyze, interpret & undertake data analysis
- Report conclusions from their analyses

- Database design, Production coding, Data collection, data warehousing, Data transformation, Work diligently with data

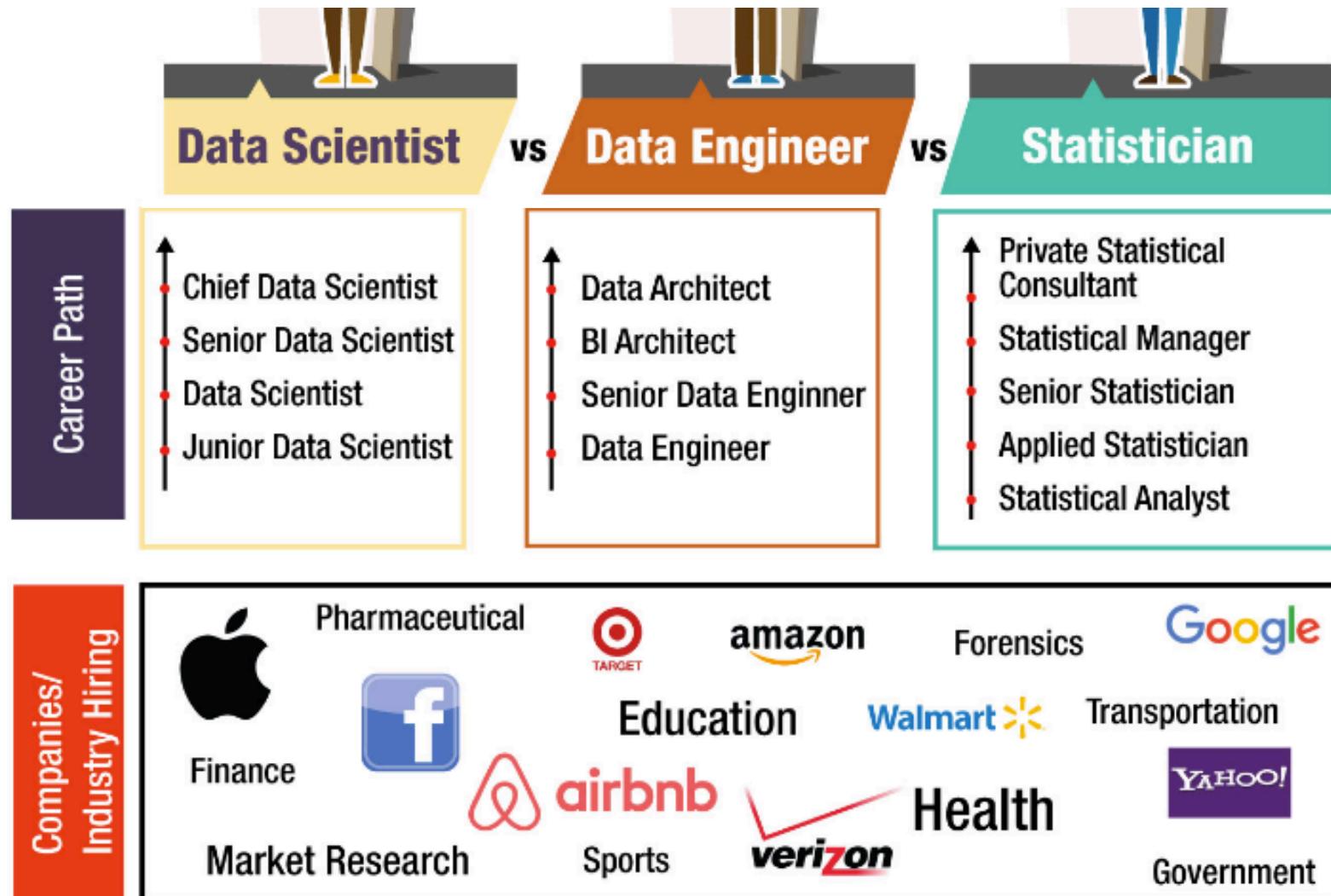
- Technical and Analytics Skills, Mathematics, Operational Research, Writing skills, Ability to Analyze, Model and interpret data, Flair of explaining difficult concepts in simple manner

source: <http://www.analyticsvidhya.com/blog/2015/10/job-comparison-data-scientist-data-engineer-statistician/>

Data Scientist vs Data Engineer vs Statistician



source: <http://www.analyticsvidhya.com/blog/2015/10/job-comparison-data-scientist-data-engineer-statistician/>



source: <http://www.analyticsvidhya.com/blog/2015/10/job-comparison-data-scientist-data-engineer-statistician/>



COMET

the real **struggle**

okay so ML is here but what's stopping us?

our brains are so **complex** they're
too difficult to understand. but
they're powerful like that.



www.shutterstock.com - 325209851



COMET

→ planes **don't flap their wings** to fly.
but birds don't take off from trees
either. a car can run faster than a
cheetah. but a car can **never climb** a
tree.

-anonymous

TWEET TO ME @jordandoinwork

EMAIL ME jordan.deja@dlsu.edu.ph

VISIT ME <http://spark.adobe.com/page/A1gpP/>

Thank you!

twitter.com/jordandoinwork

jordan.deja@dlsu.edu.ph