

Catch me if you can

How to fight Fraud, Waste and Abuse using Machine Learning AND Machine Teaching?

Cupid Chan 2020-11-12

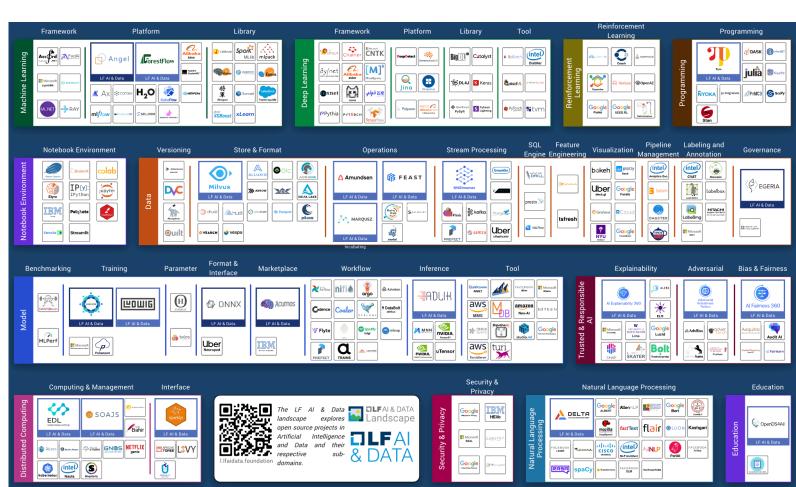






+ JLFA







Message

More...

Jacqueline Mars · 1st in



Miss Hall's School

Anthropology at Miss Hall's School

Washington D.C. Metro Area · 88 connections · Contact info

Highlights



16 mutual connections

You and Jacqueline both know Chung Lau, Siva Natarajan, Ph.D., and 14 others

Experience



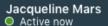
Co-Founder AMERICAN CANDY CO LIMITED

Sep 1982 - Sep 2001 · 19 yrs 1 mo

Education



Miss Hall's School Anthropology









Jacqueline Mars · 1st Anthropology at Miss Hall's School

TODAY



Jacqueline Mars • 6:58 PM

Hello 👋



Jacqueline Mars • 8:32 PM

Hi Cupid, Good to view your network, i am an investor, I always honored and inspired to meet great profile and great men like you, i am hoping to see how well we can work together in future has i have just activated my LinkedIn! Jacqueline



Cupid Chan • 9:02 PM

Welcome to LinkedIn, Jacqueline! Let's keep in touch and see what can we work together in the future.



Jacqueline Mars • 10:02 PM

Absolutely

I am an investor and an anthropologist.i am from Washington DC but presently based in The plains, Virginia. i am also founders of the American candy company Mars, Incorporated.

www.mars.com



Cupid Chan • 10:13 PM

I am in a Virginian too. Nice to meet you!



Jacqueline Mars • 10:15 PM

Same here

What do you do?



Cupid Chan • 10:18 PM

I help organization to leverage the data they have using AI and Analytics.

For example, use in prevent Fraud, Waste and Abuse, predicting sales... etc.



Jacqueline Mars • 10:19 PM Interesting



Jacqueline Mars • 10:21 PM

I would like you to share with me in exclusive about yourself and what you do in your industry, what potentials you have. And how profitable it is..as I am seeking to invest in private organizations, Government organizations and even individuals..This is why I am a business woman and rated world wide.



Cupid Chan • 10:25 PM

By "exclusive", you mean in what sense?



Jacqueline Mars • 10:26 PM Yes, exactly



Cupid Chan • 10:29 PM

I am sorry, but exactly what?

LinkedIn Story to be continued...



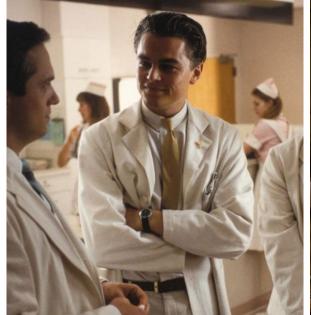


10 seconds Polling Question

Have you watched the movie "Catch me if you can"?











How to fight
Fraud, Waste
and Abuse using
Machine
Learning AND
Machine
Teaching?



Presented by Cupid Chan



WORLD USABILITY DAY

Fraud, Waste, Abuse and Error



Intentional Deception



AbuseBending the rules



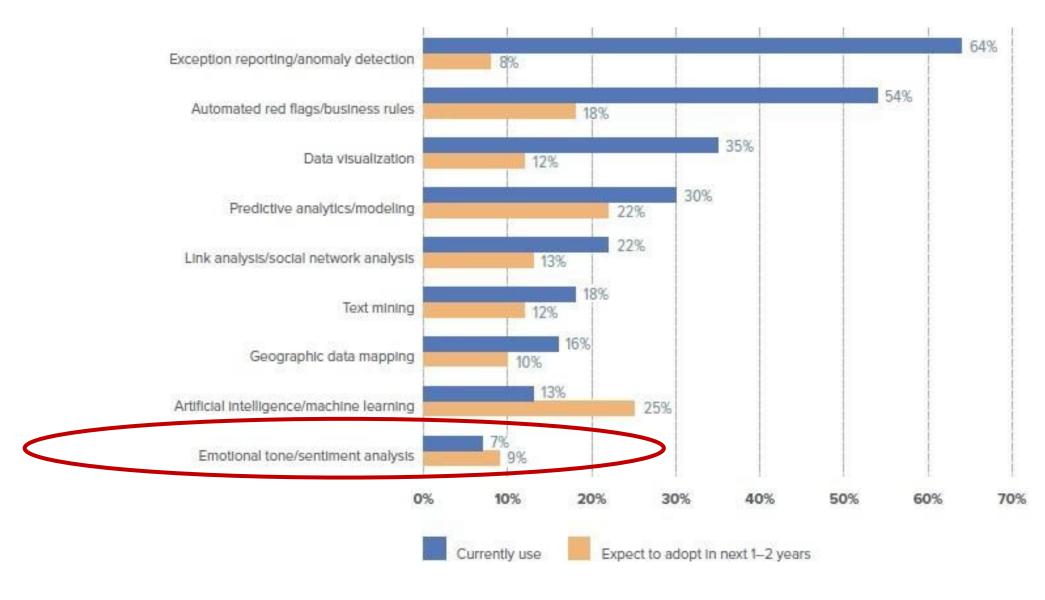
Waste Inefficiency



ErrorUnintentional fault



FIG. 1 What data analysis techniques do organizations use to fight fraud?



10 seconds Polling Question - 1

Which of the following popular fraud costing the most?

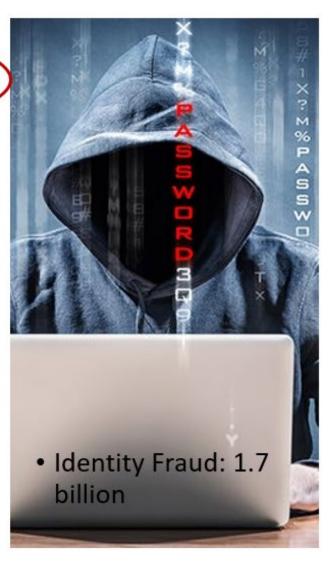
- Credit card Fraud
- Healthcare Fraud
- Identity Fraud



Cost of Fraud







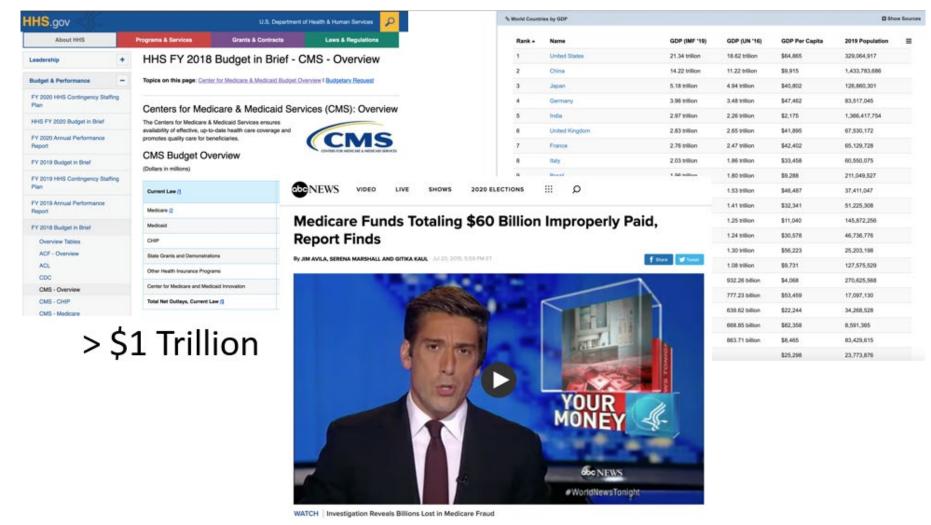
For more information about the cost of fraud, click the links below.

https://www.bcbsm.com/health-care-fraud/fraud-statistics.html

https://losspreventionmedia.com/credit-card-fraud-news-2018-update/

https://www.javelinstrategy.com/coverage-area/2019-identity-fraud-report-fraudsters-seek-new-targets-and-victims-bear-brunt

CMS Costs of Fraud



For more information about the costs of fraud, click the links below.

http://worldpopulationreview.com/countries/countries-by-gdp/

https://www.hhs.gov/about/budget/fy2018/budget-in-brief/cms/index.html#overview

http://abcnews.go.com/Politics/medicare-funds-totaling-60-billion-improperly-paid-report/story?id=32604330



Medicare Provider Utilization and Payment Data: Physician and Other Supplier



ne Abou	t CMS Newsroom Archive	Share	🕗 Help ᇦ Print
	type search term here		Search

Medicare

Medicaid/CHIP

Medicare-Medicaid Coordination Private Insurance Innovation Center Regulations & Guidance Research, Statistics, Data & Systems Outreach & Education

Home > Research, Statistics, Data and Systems > Medicare Provider Utilization and Payment Data > Medicare Provider Utilization and Payment Data; Physician and Other Supplier

Medicare Provider Utilization and Payment Data

Medicare Provider Utilization and Payment Data: Physician and Other Supplier

Medicare Provider Utilization and Payment Data: Inpatient

Medicare Provider Utilization and Payment Data: Outpatient

Medicare Provider Utilization and Payment Data: Part D Prescriber

Medicare Provider Utilization and Payment Data: Referring Durable Medical Equipment, Prosthetics, Orthotics and Supplies

Medicare Provider Utilization and Payment Data: Post-Acute Care and Hospice

Legacy Medicare Provider
Utilization and Payment Data:
Home Health Agencies

<u>Legacy Medicare Provider</u>
<u>Utilization and Payment Data:</u>
Skilled Nursing Facilities

<u>Legacy Medicare Provider</u>
<u>Utilization and Payment Data:</u>
Hospice Providers

Public Comment on the Release of Medicare Physician Data

Medicare Provider Utilization and Payment Data: Physician and Other Supplier

The Physician and Other Supplier Public Use File (Physician and Other Supplier PUF) provides information on services and procedures provided to Medicare beneficiaries by physicians and other healthcare professionals. The Physician and Other Supplier PUF contains information on utilization, payment (allowed amount and Medicare payment), and submitted charges organized by National Provider Identifier (NPI), Healthcare Common Procedure Coding System (HCPCS) code, and place of service. This PUF is based on information from CMS administrative claims data for Medicare beneficiaries enrolled in the fee-for-service program. The data in the Physician and Other Supplier PUF covers calendar years 2012 through 2017 and contains 100% final-action physician/supplier Part B non-institutional line items for the Medicare fee-for-service population.

While the Physician and Other Supplier PUF has a wealth of information on payment and utilization for Medicare Part B services, the dataset has a number of limitations. Of particular importance is the fact that the data may not be representative of a physician's entire practice as it only includes information on Medicare fee-for-service beneficiaries. In addition, the data are not intended to indicate the quality of care provided and are not risk-adjusted to account for differences in underlying severity of disease of patient populations. For additional limitations, please review the methodology document available below.

Medicare Physician and Other Supplier Data CY 2017 Medicare Physician and Other Supplier Data CY 2016

Medicare Physician and Other Supplier Data CY 2015

Medicare Physician and Other Supplier Data CY 2014

Medicare Physician and Other Supplier Data CY 2013
Medicare Physician and Other Supplier Data CY 2012

Inquiries regarding this data can be sent to MedicareProviderData@cms.hhs.gov.

To receive email notifications, please sign up for the Medicare Provider Data GovDelivery subscription here.

Downloads

Medicare Physician and Other Supplier PUF Methodology [PDF, 357KB] Medicare Physician and Other Supplier PUF Frequently Asked Questions [PDF, 135KB]

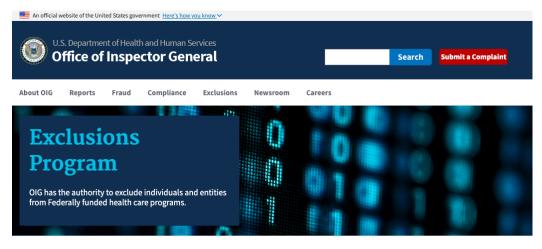
Page last Modified: 07/12/2019 10:23 AM

2012	2013	2014	2015	2016	2017
# NPI					
ABC NPPES_PROVIDER_LAST_ORG_NAME	ABC NPPES_PROVIDER_LAST_ORG_NAME	ABC nppes_provider_last_org_name	ABC nppes_provider_last_org_name	RBC NPPES_PROVIDER_LAST_ORG_NAME	ABC nppes_provider_last_org_name
RBC NPPES_PROVIDER_FIRST_NAME					
ABC NPPES_PROVIDER_MI					
RBC NPPES_CREDENTIALS	RBC NPPES_CREDENTIALS	RBC nppes_credentials	ABC nppes_credentials	RBC NPPES_CREDENTIALS	RBC nppes_credentials
NPPES_PROVIDER_GENDER	NPPES_PROVIDER_GENDER	nppes_provider_gender	nppes_provider_gender	NPPES_PROVIDER_GENDER	nppes_provider_gender
RBC NPPES_ENTITY_CODE	RBC NPPES_ENTITY_CODE	ABC nppes_entity_code	RBC nppes_entity_code	RBC NPPES_ENTITY_CODE	RBC nppes_entity_code
ABC NPPES_PROVIDER_STREET1	RBC NPPES_PROVIDER_STREET1	ABC nppes_provider_street1	ABC nppes_provider_street1	ABC NPPES_PROVIDER_STREET1	ABC nppes_provider_street1
RBC NPPES_PROVIDER_STREET2					
ABC NPPES_PROVIDER_CITY	RBC NPPES_PROVIDER_CITY	ABC nppes_provider_city	RBC nppes_provider_city	ABC NPPES_PROVIDER_CITY	ABC nppes_provider_city
# NPPES_PROVIDER_ZIP					
NPPES_PROVIDER_STATE	■ NPPES_PROVIDER_STATE	nppes_provider_state	nppes_provider_state	NPPES_PROVIDER_STATE	nppes_provider_state
RBC NPPES_PROVIDER_COUNTRY					
ABC PROVIDER_TYPE	RBC provider_type				
 MEDICARE_PARTICIPATION_INDICATOR 					
PLACE_OF_SERVICE	 PLACE_OF_SERVICE 	place_of_service	place_of_service	RBC PLACE_OF_SERVICE	place_of_service
# HCPCS_CODE					
ABC HCPCS_DESCRIPTION					
 HCPCS_DRUG_INDICATOR 					
# LINE_SRVC_CNT					
# BENE_UNIQUE_CNT					
# BENE_DAY_SRVC_CNT					
#,# AVERAGE_MEDICARE_ALLOWED_AMT	## AVERAGE_MEDICARE_ALLOWED_AMT	## average_Medicare_allowed_amt	## average_Medicare_allowed_amt	## AVERAGE_MEDICARE_ALLOWED_AMT	## average_Medicare_allowed_amt
# STDEV_MEDICARE_ALLOWED_AMT	# STDEV_MEDICARE_ALLOWED_AMT	## average_submitted_chrg_amt	# average_submitted_chrg_amt	# AVERAGE_SUBMITTED_CHRG_AMT	#,# average_submitted_chrg_amt
# AVERAGE_SUBMITTED_CHRG_AMT	## AVERAGE_SUBMITTED_CHRG_AMT	## average_Medicare_payment_amt	## average_Medicare_payment_amt	## AVERAGE_MEDICARE_PAYMENT_AMT	## average_Medicare_payment_amt
# STDEV_SUBMITTED_CHRG_AMT	## STDEV_SUBMITTED_CHRG_AMT	#,# average_Medicare_standard_amt	## average_Medicare_standard_amt	## AVERAGE_MEDICARE_STANDARD_AMT	#,# average_Medicare_standard_amt
## AVERAGE_MEDICARE_PAYMENT_AMT	## AVERAGE_MEDICARE_PAYMENT_AMT				

STDEV_MEDICARE_PAYMENT_AMT

STDEV_MEDICARE_PAYMENT_AMT

List of Excluded Individuals and Entities (LEIE)



Exclusions Program Online Searchable Database LEIE Downloadable Databases Monthly Supplement This webpage provides the authority to exclud programs for a variety Those that are exclude items or services they benefits funded directly benefits funded directly the services they benefit for the services the services they benefit for the services the services the services they benefit for the services the services

Archive

Quick Tips

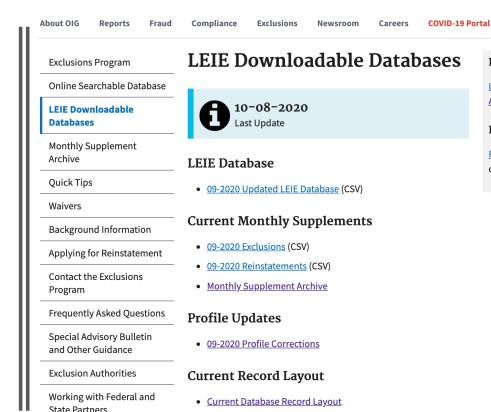
Waivers

Background Information

Applying for Reinstatement

This webpage provides information about OIG's exclusion authority and activities. OIG has the authority to exclude individuals and entities from Federally funded health care programs for a variety of reasons, including a conviction for Medicare or Medicaid fraud. Those that are excluded can receive no payment from Federal healthcare programs for any items or services they furnish, order, or prescribe. This includes those that provide health benefits funded directly or indirectly by the United States (other than the Federal Employees Health Benefits Plan).

OIG maintains a list of all currently excluded individuals and entities called the <u>List of Excluded Individuals/Entities</u> (LEIE). Anyone who hires an individual or entity on the LEIE may be subject to civil monetary penalties (CMP). To avoid CMP liability, health care entities should routinely check the list to ensure that new hires and current employees are not on it



Related Information

About the LEIE Files.

Instructions and information

File-Type Questions?

Frequently Asked Questions

concerning the CSV file type.

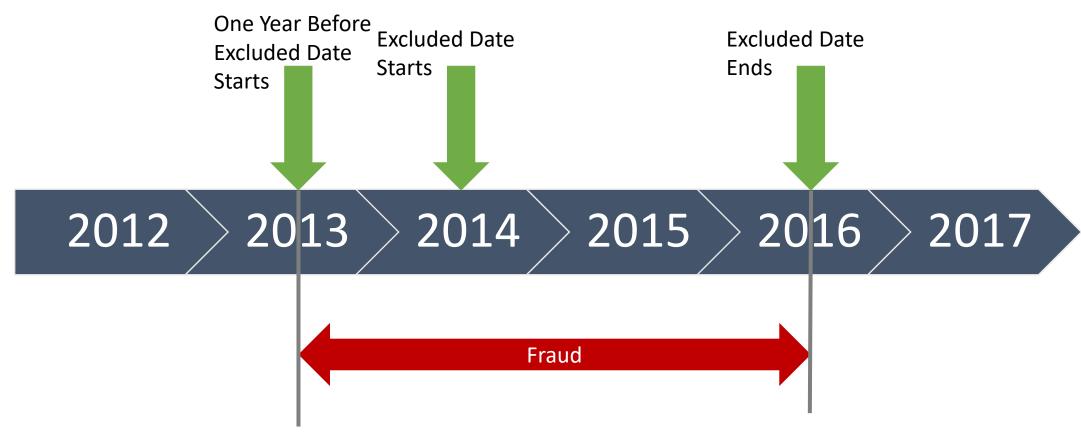
ABC	LASTNAME
ABC	FIRSTNAME
ABC	MIDNAME
ABC	BUSNAME
ABC	GENERAL
ABC	SPECIALTY
ABC	UPIN
ABC	NPI
ABC	DOB
ABC	ADDRESS
ABC	CITY
P	STATE
#	ZIP
ABC	EXCLTYPE
()	EXCLDATE
ABC	
	REINDATE
ABC	

Mandatory Exclusions

Social Security Act	42 USC §	Amendment
1128(a)(1)	1320a-7(a) (1)	Conviction of program-related crimes. Minimum Period: 5 years
1128(a)(2)	1320a-7(a) (2)	Conviction relating to patient abuse or neglect. Minimum Period: 5 years
1128(a)(3)	1320a-7(a) (3)	Felony conviction relating to health care fraud. Minimum Period: 5 years
1128(a)(4)	1320a-7(a) (4)	Felony conviction relating to controlled substance. Minimum Period: 5 years
1128(c)(3) (G)(i)	1320a-7(c) (3)(G)(i)	Conviction of second mandatory exclusion offense. Minimum Period: 10 years
1128(c)(3) (G)(ii)	1320a-7(c) (3)(G)(ii)	Conviction of third or more mandatory exclusion offenses. Permanent Exclusion

See <u>Office of Inspector General Exclusion Authorities</u> at https://oig.hhs.gov/exclusions/authorities.asp

Define what is fraud based on the data set



All HCPCS within the period are considered fraud for the NPI



Overall data ingestion flow



Final data set structure

ABC	provider_type	Provider's specialty, e.g. Internal Medicine, Dermatology
•	nppes_provider_gender	Provider Gender
ABC	hcpcs_code	Procedure or Service performed by the provider
#	line_srvc_cnt	Number of procedures or services the provider performed
#	bene_unique_cnt	Number of distinct Medicare beneficiaries receiving the service/procedure
#	bene_day_srvc_cnt	Number of distinct Medicare beneficiaries per day by the provider
#,#	average_submitted_chrg_amt	Average charge the provider submitted for the service or procedure
#,#	average_medicare_payment_amt	Average payment made to a provider per claim for the service
•	fraud	Fraud label based on the logic described before



Let's predict using Machine Learning!

Based on my rich experience in Al (3), I can build a model guaranteed with 99.9% accuracy within 10 seconds!

EVERYTHING Is NOT Fraud



Confusion Matrix

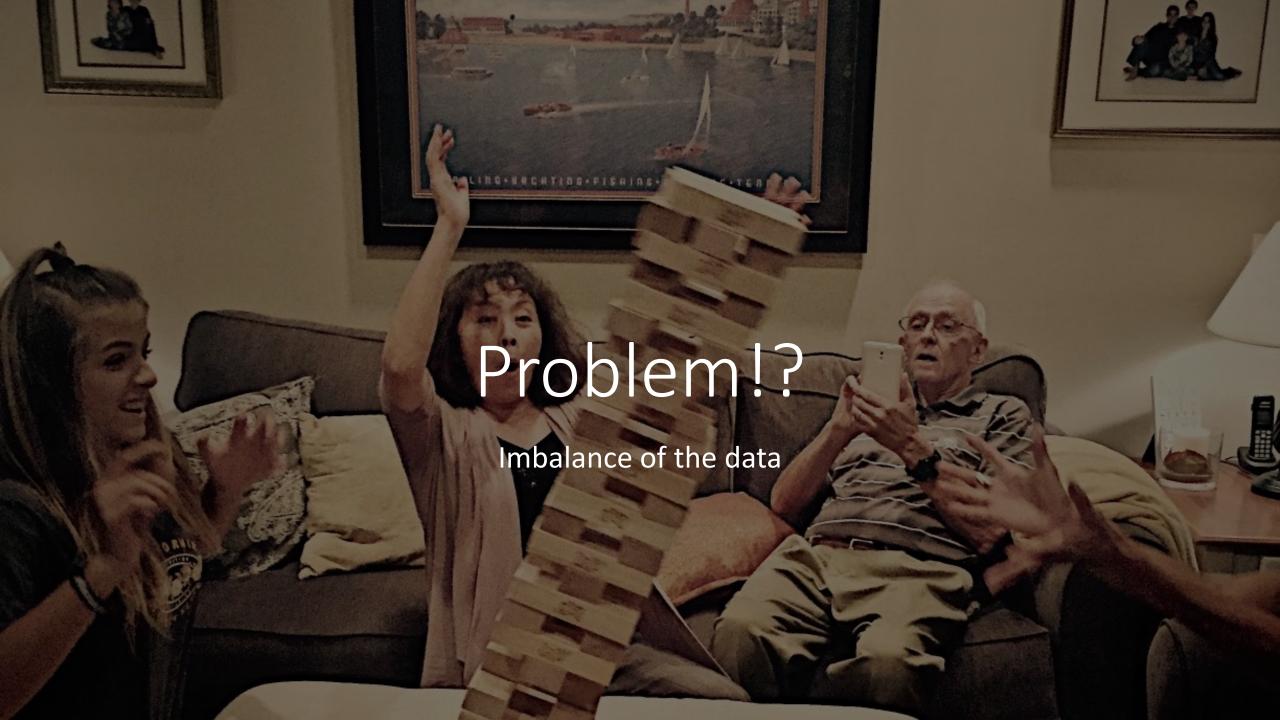
• Total: 54,337,938

• Normal: 54,333,245

• Fraud: 4,693 (0.0086%)

		Predicted Class		
		Positive	Negative	
Actual Class	Positive	True Positive (TP)	False Positive (FN)	Sensitivity/Recall TP $\overline{TP + FN}$
	Negative	False Negative (FP)	True Negative (TN)	Specificity $\frac{TN}{TN + FP}$
		Precision $\frac{TP}{TP + FP}$	Negative Predictive Value TN	Accuracy $\frac{TP + TN}{TP + TN + FP + FN}$





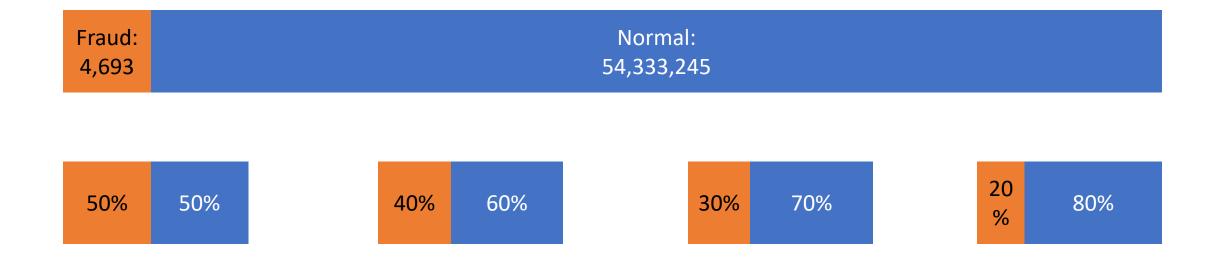


- Credit Card Fraud
- Manufacturing Defect
- Rare Disease Diagnosis
- Natural Disasters

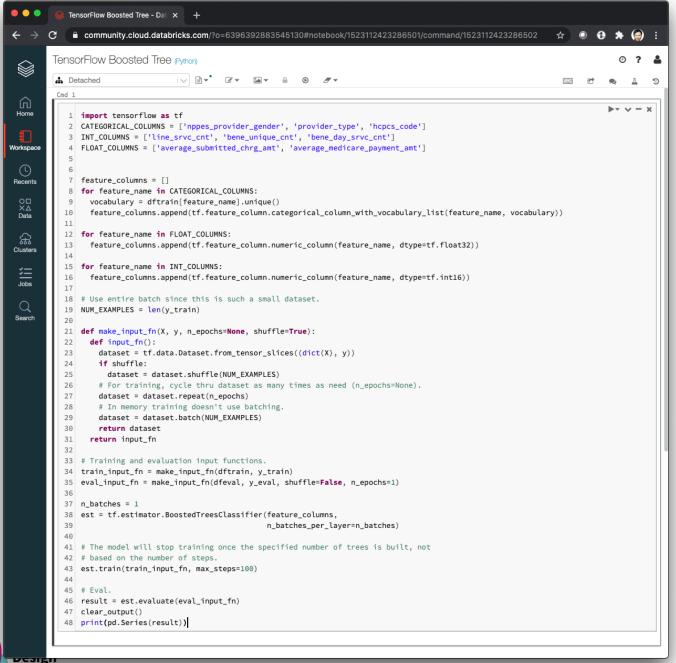




Random Under Sampling (RUS)







Using 50:50 Class Distribution

TensorFlow Boosted
Tree Classifier

Accuracy: 0.760789

Precision: 0.706313

Recall: 0.857778

AUC: 0.845611

Potential Improvement

- Add back Geographical information to the data set in analysis
- Add beneficiary data to form a graph analysis. Right now we only analyze from Provide side
- More granular e.g. by type
- Add more metrics (Medicare Standard Amount, Medicare Allowed Amount)
- A lot of missing NPI in LEIE. looking up missing NPI numbers in the National Plan and Provider Enumeration System (NPPES) registry



Even with those improvements, there are still limitations

- Tagged data not always be available
- Not good for emerging anomalies with entirely new and more sophisticated forms





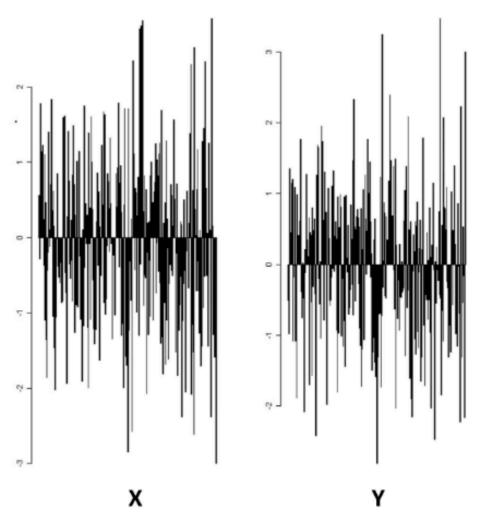
Unsupervised Learning

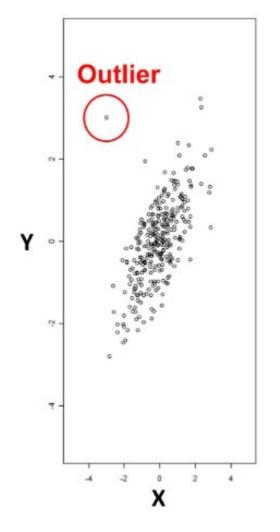
 Good for detecting Outlier, but it doesn't mean that is Fraud.

• It provides hint to start finding Fraud.



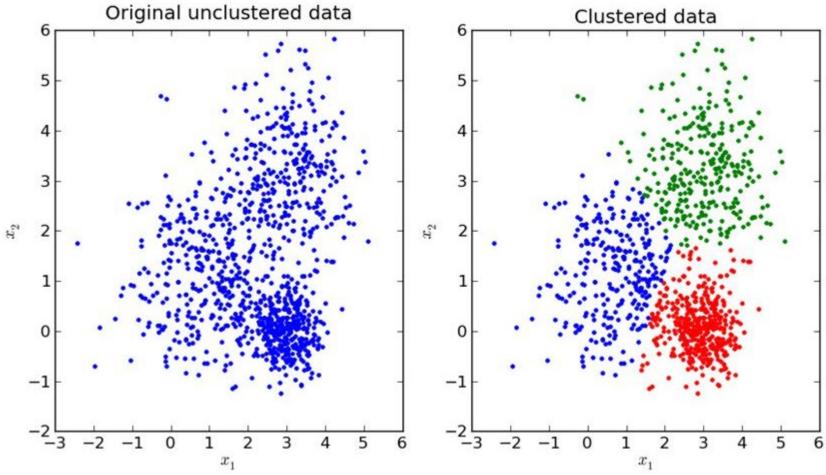
Where is the Outlier?







Unsupervised Clustering



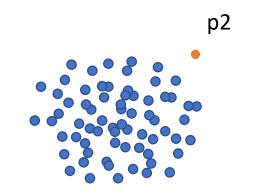


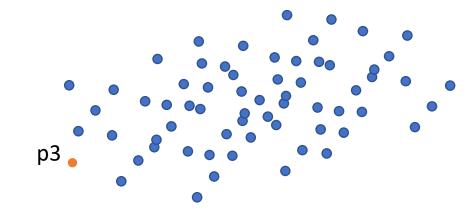


10 seconds Polling Question: Which point is considered outlier?

Nearest Neighbor Approach

- By distance
- Having the largest distance away from closest points
- Only p1 is considered outlier





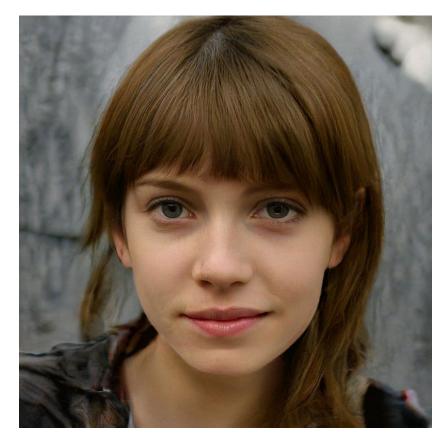
Density based Approach

- By density
- Having the lowest density among closest points
- Both p1 and p2 are considered outlier



10 seconds Polling Question - 3

You must know her for this 3rd approach



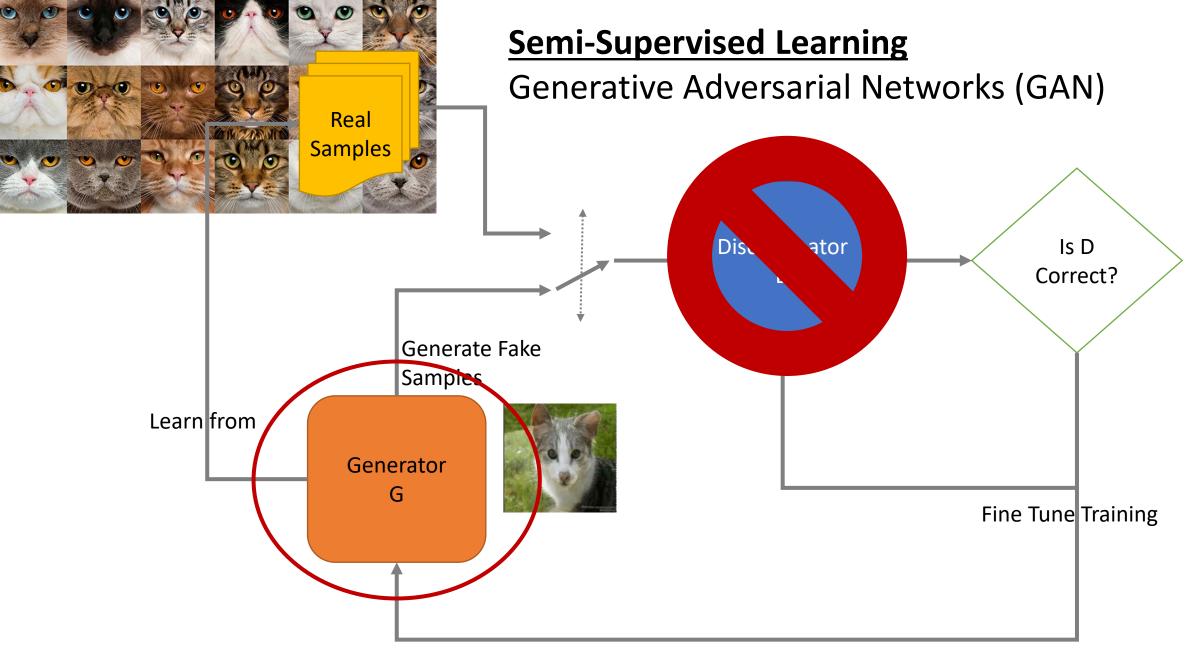
Human-Centered Design Center of Excellence

Who is she?

- She committed a \$100M fraud in a European bank last year by guessing the admin password using AI algorithm
- She is actually a man dressed in disguise to fool the airport security smuggling 500 fake passports (with valid passport numbers produced by AI) in US last year
- The person inventing this 3rd approach
 Al algorithm

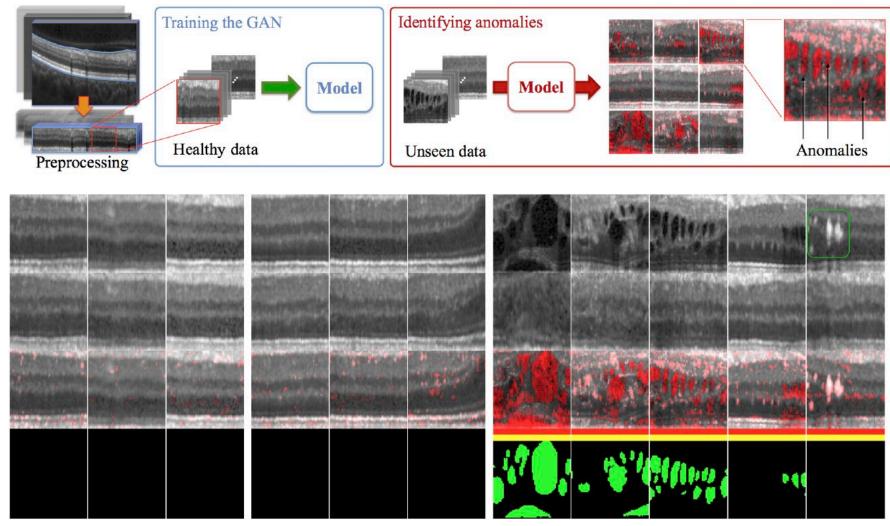
For more information about stylegan, click the links below

http://stylegan.xyz/paper https://github.com/NVlabs/stylegan





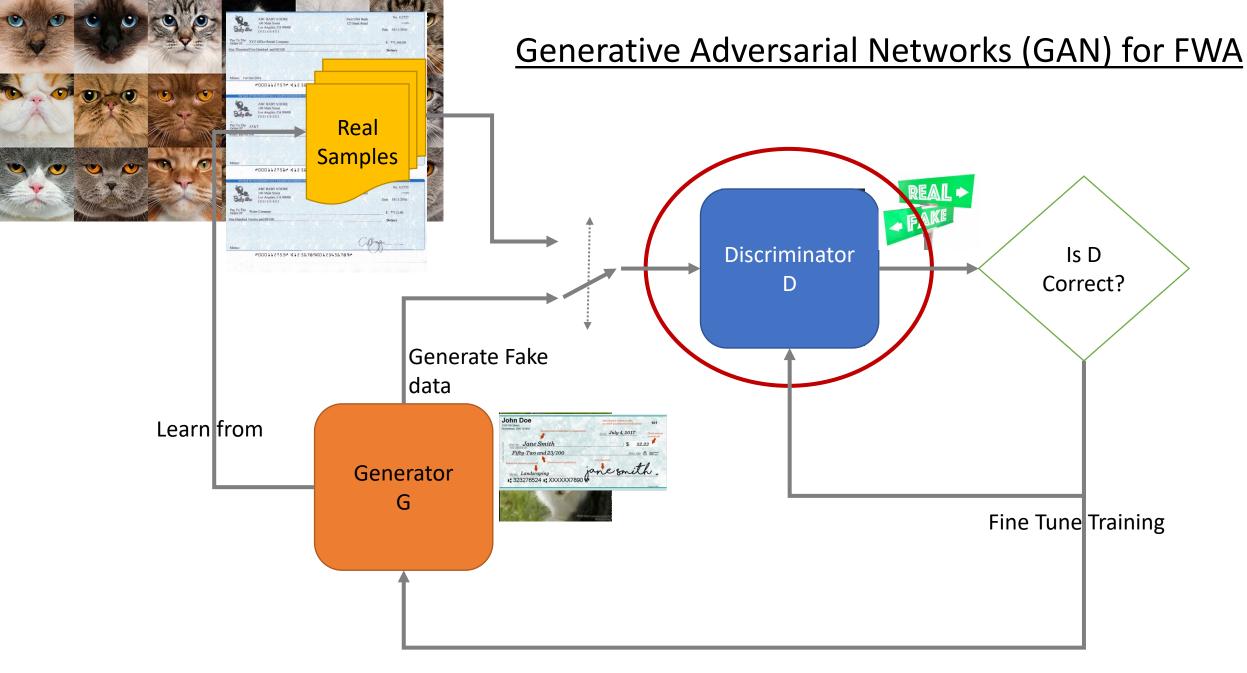
AnoGAN



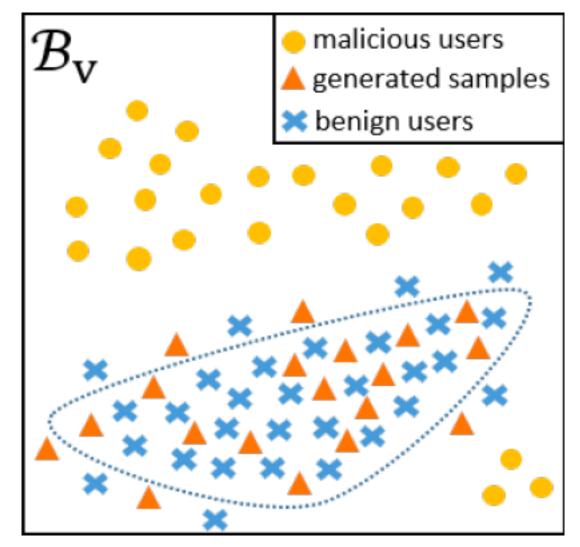


Unsupervised Anomaly Detection with Generative Adversarial Networks to Guide Marker Discovery: Thomas Schlegl, Philipp Seeb®ock, Sebastian M. Waldstein, Ursula Schmidt-Erfurth, and Georg Langs (Medical University Vienna, Austria) CCSQ WORLD USABILITY DAY http://arxiv.org/abs/1703.05921





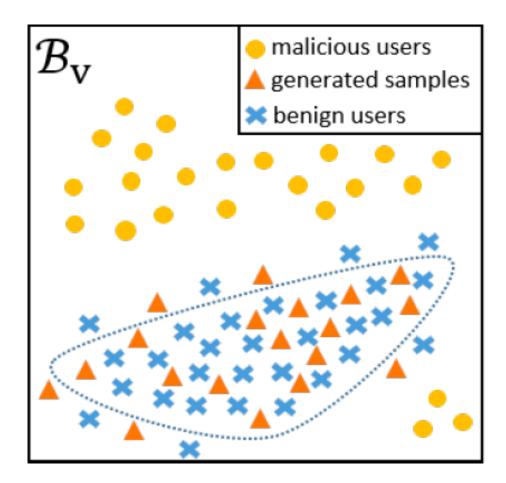
Traditional GAN

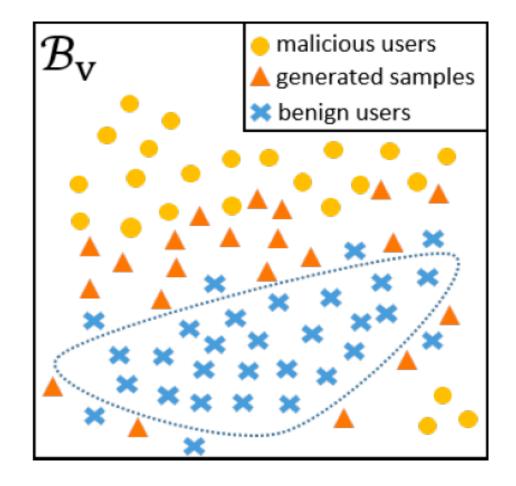


See "One-Class Adversarial Nets for Fraud Detection." Click the link below. https://arxiv.org/pdf/1803.01798.pdf



One-Class Adversarial Nets (OCAN) GAN





(a) Regular GAN

(b) Complementary GAN



Advantage of One-Class Adversarial Nets

- No need for fraud data
 - No need to manually prepare a mixed training data set, which is usually has a very few fraud data to start with
- Discriminator will take in either real benign or generated malicious
 - More adaptive to different kinds of malicious behavior
- Adapt to newly emerged normal user pattern



Recap: What we have talked about so far...

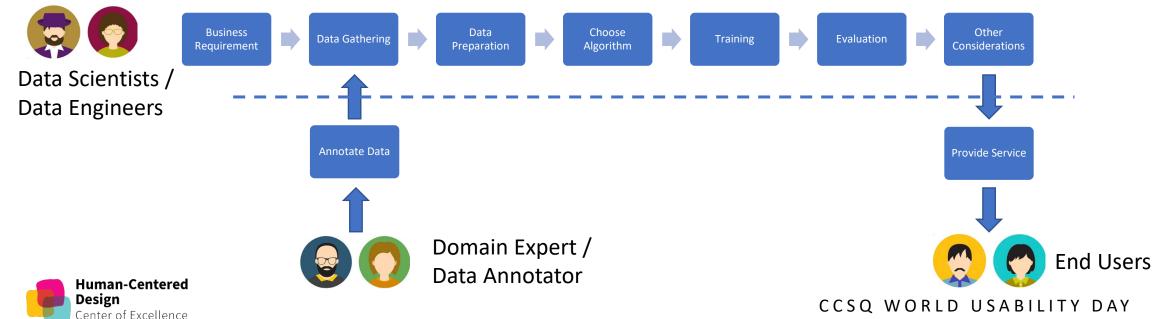




Al Recipe for Data Kitchen (for Machine Learning)



https://www.linkedin.com/pulse/airon-chef-cook-up-ai-your-data-kitchen-cupid-chan/https://www.youtube.com/watch?v=xifaCq4F7bg



Human-Centered AI – Machine Learning and Teaching

Machine learns from data

Teaching in Input:

annotating correct data

Business Requirement

Data Gathering

Preparation

Data Preparation

Preparation

Training

Training

Preparation

Data Preparation

Preparation

Algorithm

Preparation

Algorithm

Data Choose Algorithm

Algorithm

Algorithm

Data Preparation

Data Prepa

I know a cat has pointy ears but found something I am not sure.

Thank you! I will update my model to recognize this type of dog



the data, and

Oh, that's called Chihuahua. It's a dog, not cat.

pattern this year, especially the number of patient has jumped 25% more then the 5 years running average.

Dr XXX has unusual

claim (potential fraud)

Teaching in Output: making appropriate decision

Machine makes decision based on the model learnt and provides service

Dr XXX's colleague
Dr YYY has had a
maternity leave for 9
months, so Dr XXX
picked up her
patients in the
interim

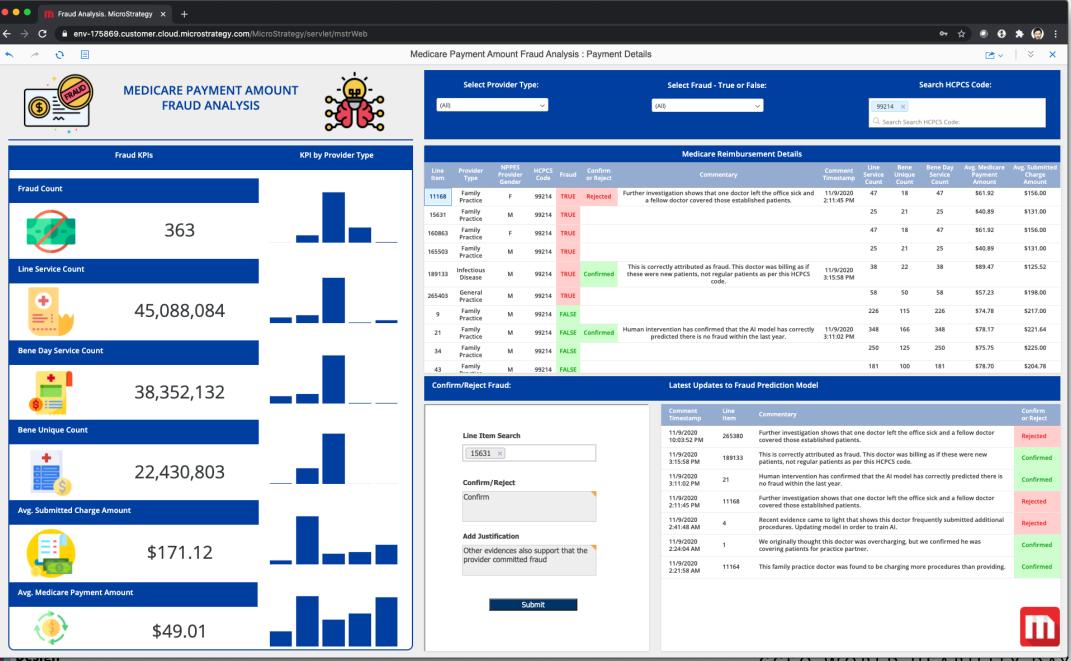
CCSQ WORLD USABILITY DAY

Collaborate and

decision making

teach in

Service consumed by human



When you talk about AI, think of IA!

Artificial Intelligence VS Intelligence Augmentation



AI + BI = CI

AI (Artificial Intelligence) — Excellence in learning with speed

BI (Business Intelligence) – Historically proven to enable human intuited the direction

CI – (Cognitive Intelligence) -Intelligence by combining the Speed of how a Machine Learn and Direction Intuited from Human Insight



2017 CONFERENCE ON HEALTH IT AND ANALYTICS (CHITA)



The Conference on Health IT and Analytics (previously known as the Workshop on Health IT & Economics) is an annual health IT and analytics research summit, including a doctoral consortium that each year gathers prominent scholars from more than 40 research institutes, and leading policy and practitioner attendees in a vibrant setting to discuss opportunities and challenges in the design, implementation and management of health information technology and analytics. Its goal is to deepen our understanding of strategy, policy and systems fostering health IT and analytics effective use and to stimulate new ideas with both policy and business implications.

This forum provides a productive venue to facilitate interaction and collaboration among academia, government, and industry. Now in its 8th year, each year CHITA draws over 100 participants.

Hosted by the Center for Health Information & Decision Systems (CHIDS), support for CHITA is provided by the Robert H. Smith School of Business and the University of Michigan School of Public Health.

We hope that you will join us for this engaging, stimulating and fun event!

FEATURED SPEAKERS



Martin "Marty" Makary
Chief, Islet Transplant
Surgery, Professor, NYT
best-selling author
Johns Hopkins University
School of Medicine



Warren D. D'Souza
Vice President, Enterprise
Data and Analytics
University of Maryland
Medical System

FEATURED PANELISTS



Cupid Chan
Chief Technology Officer
Index Analytics



Rema Padman
Professor of Management
Science and Healthcare
Informatics
Carnegie Mellon University



Anna T. Fernandez
Senior Associate, Health
Informatics
Booz Allen Hamilton



Charles Gabrial
Analytics & Data Science
Manager
Defense Health Agency



Hoo Chang Shin Solutions Architect NVIDIA



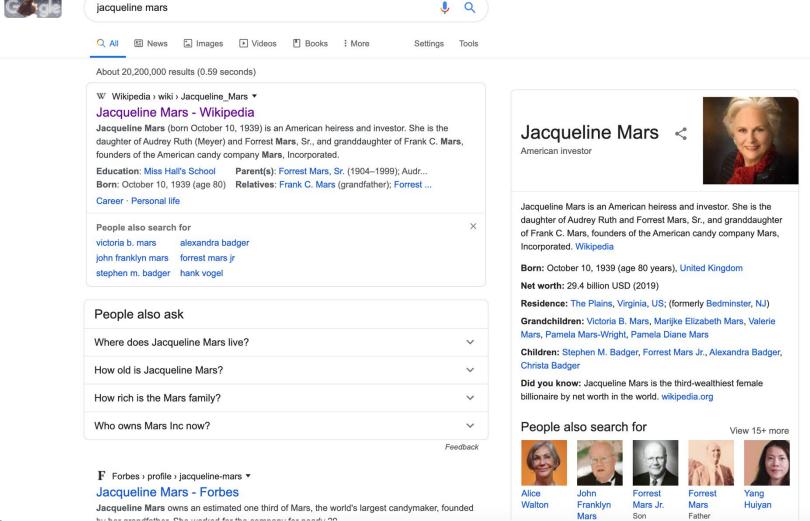
Hong Yu
Professor of
Quantitative Health
Sciences

... LinkedIn Story continues





I decided to find my most knowledgeable friend for help







Cupid Chan • 10:29 PM

I am sorry, but exactly what?



Jacqueline Mars • 10:31 PM

What you do in your place of work and the profitable it is.



Cupid Chan • 10:31 PM

i see

Cupid Chan is typing...



Jacqueline Mars • 10:38 PM

Don't you want to talk to me about it's fine I'm sorry (2)





Cupid Chan • 10:38 PM

Cupid Chan is a seasoned professional who is well-established in the industry. His journey started out as one of the key players in building a world-class BI platform. Aside from holding various technical accreditations his credential extends into husiness



Jacqueline Mars • 10:42 PM

WOW

Cool

The real: Frank Abagnale

Cupid Chan

- Board of Directors and Technical Steering Committee, Chairperson of BI & Al Project, Linux Foundation ODPi
- Senior Fellow & Adjunct Professor, University of Maryland College Park

www.linkedin.com/in/cupidchan/@cupidckchan

