

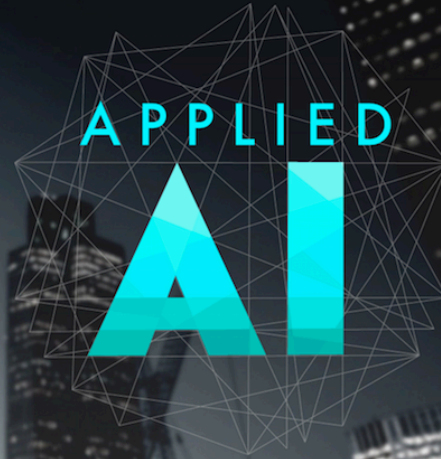
HOW IS DATA SCIENCE GOING TO IMPROVE INSURANCE?

MICHAEL CRAWFORD &
JONATHAN SEDAR

APPLIED AI LTD

@APPLIED_AI

APPLIED.AI





HOW ARE WE *ALREADY* USING
DATA SCIENCE TO IMPROVE
INSURANCE?

THREE STORIES FROM OUR EXPERIENCES OVER THE PAST TWO YEARS.

-

1. CURATING EXTERNAL DATA TO BETTER
UNDERSTAND CUSTOMERS
2. BRINGING STATISTICAL MODELING INTO THE
BOARDROOM
3. APPLYING MORE EXOTIC MACHINE LEARNING TO THE
BUSINESS



APPLIED AI IS A DATA SCIENCE
CONSULTANCY

WE HELP LEADING INSURERS CREATE A COMPETITIVE
ADVANTAGE THROUGH **APPLIED ARTIFICIAL
INTELLIGENCE**

WHO ARE WE?

A SMALL BUT EXPERIENCED TEAM WITH EXPERTISE IN
STATISTICS, MACHINE LEARNING, ACTUARIAL SCIENCE,
SOFTWARE DEVELOPMENT, FINANCIAL SYSTEMS AND
CONSULTANCY.



Michael Crawford

Actuarial Science | Financial Systems | Software
Dev

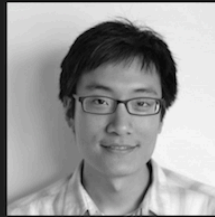
Jonathan Sedar

Machine Learning | Physical Sciences |
Consulting



A Team of Expert Practitioners

Quantitative Finance | Statistics | Software Development |
Insurance



FOUNDING MYTH

WE MET IN THE PUB FOLLOWING AN R MEETUP
... NOT A MILLION MILES AWAY FROM THIS
WE BONDED OVER BEER & GENERAL GEEKINESS

**AND THOUGHT THAT MACHINE LEARNING &
INSURANCE WERE A NATURAL FIT**

LETS ASK INSURANCE COMPANIES

A collection of colorful bingo balls scattered on a dark surface. The balls are in various colors including red, orange, yellow, green, blue, purple, and pink. Each ball has a white circular area in the center containing a black number and the 'BINGO ansdag' logo. The numbers visible are 12, 43, 3, 9, 25, 4, 5, 2, 3, and 3. The word 'BINGO!' is written in large white letters across the center of the image.

BINGO!



SO WHAT HAPPENED
NEXT?



PART ONE

CURATING EXTERNAL DATA TO BETTER
UNDERSTAND CUSTOMERS

-

OR WHAT HAPPENED WHEN WE CUT EXPERIAN OUT
OF THE LOOP

-

CLUSTERING
INTROSPECTION
VISUALISATION

BACKGROUND

WE WORK MAINLY WITH INSURANCE COMPANIES
THEY DON'T HAVE A REPUTATION FOR BEING EXCITING
BUT FROM A DATA SCIENCE POINT OF VIEW...

A promotional image for the movie 'Indiana Jones and the Temple of Doom'. Indiana Jones is shown from the chest up, wearing his signature brown fedora and a dark, sleeveless vest over a light-colored shirt. He is holding a whip in his right hand. The background is a lush, green jungle with rolling hills. The text 'IT'S QUITE INTERESTING!' is overlaid in white, sans-serif font in the upper center. The title 'INDIANA JONES' is written in a large, stylized, yellow-to-orange gradient font with a black outline, slanted upwards. Below it, the subtitle 'and the TEMPLE OF DOOM.' is written in a smaller, yellow, sans-serif font.

IT'S QUITE
INTERESTING!

INDIANA JONES
and the
TEMPLE OF DOOM.

The
complete

RIPPING YARNS

2
DISC SET

By Michael Palin
and Terry Jones

HERE'S A STORY



THE PROBLEM

Nº 301114



The Great Eastern Life Assurance Co., Ltd.

(Incorporated in the Straits Settlements.)

HEAD OFFICE: - - SINGAPORE.

POLICY No.	ASSURED
70583	Ng Hop Choon
Renewed for 12 months from 20th December 1937	Premium \$884.50

OUR TERM
INSURANCE POLICIES
ARE LAPSING

Received this 20th day of January 1937

the above amount being the Premium due as stated, agreeably with the terms of the Policy and those endorsed on the back hereof.

This Policy is not valid nor operative unless this receipt is countersigned by the Agent of the Company when the said premium is paid.

Horace W. Raper

John A. ... Agent at Head Office, Managing Director.

Where the age of the assured has not been proved, it is very desirable that satisfactory evidence of age be produced to the Company, so that the age may be admitted during the lifetime of the Policy-holder.





The Great Eastern Life Assurance Co., Ltd.

TERM INSURANCE

Extra Interesting!

POLICY NO.	ASSURED
<p>WHEN AN INSURANCE COMPANY SETS A POLICY UP:</p> <ul style="list-style-type: none"> • IT PAYS A COMMISSION TO THE BROKER WHO SOLD THE POLICY 	
<p>18 months from 20th Dec</p> <ul style="list-style-type: none"> • IT MAY HAVE TO SEND YOU FOR A MEDICAL - AND PAY 	<p>Premium \$884.50</p>

Received the above amount of premium due as stated, agreeably with the terms of the Policy and those endorsed on the back hereof.

January 1907

This Policy is not valid nor operative unless this receipt is countersigned by the Agent of the Company when the said premium is paid.

IT'S 2 YEARS BEFORE THE INSURER IS IN THE BLACK
 SO THEY REALLY WANT YOU TO STICK AROUND

Harold W. Rafer Director.

Where the age of the assured has not been proved, it is very desirable that satisfactory evidence of age be produced to the Company, so that the age may be admitted during the lifetime of the Policy-holder.

THE THING IS

Still Really Interesting!

PEOPLE DON'T WANT TO STICK AROUND!

- IN THE RECESSION THEY WERE DROPPING LIKE FLIES
- WE WERE ASKED IF WE COULD FIGURE OUT WHY
- AND TRY FIND WAYS TO REDUCE IT
 - We used survival analysis (of which more later)
 - After a few weeks we had a good model

ALONG THE WAY WE NOTICED SOMETHING

A row of modern, two-story houses with stone and white siding and dark roofs. In the foreground, there is a large, desolate area covered in rubble, debris, and overgrown grass, with a small pond on the left. A person is visible in the distance, walking through the debris. The sky is overcast and grey.

THE ATTRITION RATE
WAS SKY-HIGH IN
NEW ESTATES



GEOGRAPHIC EFFECTS

A black and white photograph of three men standing side-by-side against a plain, light-colored background. The man on the left is dressed in a dark, well-tailored suit with a white shirt and a dark tie, wearing a dark bowler hat. The man in the center is wearing a dark suit with a white shirt and a dark tie, wearing a dark fedora hat. The man on the right is wearing a light-colored, possibly white, long-sleeved shirt with a dark scarf, wearing a light-colored flat cap. The text "AND SOCIOECONOMIC EFFECTS" is overlaid in the center of the image in a white, sans-serif font.

AND
SOCIOECONOMIC
EFFECTS

HOW CAN WE USE THESE EFFECTS?

- TO ENCOURAGE CUSTOMERS TO STAY
- TO HELP PRICE RISK
- IDENTIFY NEW MARKETS


WE CALLED EXPERIAN...

IT WAS A SHORT CONVERSATION

I'D LIKE SOCIOECONOMIC INFORMATION FOR 250K
ADDRESSES

...

HOW MUCH!?

A man with dark hair, wearing a grey patterned jacket, is looking directly at the camera. He is in a space station environment, with a circular hatch visible behind him. In the foreground, there is a table with a map or technical drawing on it. The text "WE CAN DO IT CHEAPER & BETTER OURSELVES" is overlaid in white, bold, sans-serif font across the center of the image.

WE CAN DO IT
CHEAPER & BETTER
OURSELVES

GEOCODING

FIRST LET'S GEOCODE OUR ADDRESSES

WE HAD TWO CHOICES:

- USE GOOGLE - WHICH YOU PAY FOR
- USE NOMINATIM - FOSS / ROLL YOUR OWN

WE TRIED BOTH:

- FOR IRELAND, GOOGLE IS BETTER
- MAINLY BECAUSE ...

A photograph of a man in a wheelchair with his arms raised in front of a large stone building. In the foreground, another man in a kilt is crawling on the grass. A green bottle lies on the grass to the right.

IRISH ADDRESSES ARE PATHOLOGICAL!

BUT WE NOW HAVE A LAT / LONG FOR EACH CLIENT

SHOPPING FOR DATA

IN IRELAND A CENSUS IS DONE EVERY FIVE YEARS

- THE AMOUNT OF INFORMATION IN IT IS ASTOUNDING
- APPROXIMATELY 700 FEATURES
- COVERING 15 THEMES

LET'S LOOK AT THE DATA AVAILABLE...

CENSUS DATA THEMES

Theme	Subject	Theme	Subject
1	Sex, Age & Migration	9	Social class
2	Ethnicity & Religion	10	Education
3	Irish Language	11	Commuting
4	Families	12	Health
5	Private Households	13	Occupation
6	Housing	14	Industries
7	Hospitals & Prisons	15	PC & Internet
8	Principal Status		

A colorful, abstract map of a city, possibly San Francisco, with various colored regions and a large blue area on the right. The map is composed of many small, irregularly shaped polygons in shades of red, orange, yellow, green, purple, and blue. The blue area on the right is a large, dark blue shape that resembles a bay or a large body of water. The text "IT GETS BETTER" is centered over the map in a black, sans-serif font.

IT GETS BETTER

A map of Ireland is shown in the background, overlaid with a grid of small, irregularly shaped areas. Each area is filled with a different color, including shades of green, purple, brown, and blue. The map is centered on the island of Ireland, with the surrounding ocean visible in dark blue.

SMALL AREA MAPS

THE SMALLEST OUTPUT AREA FOR CENSUS DATA

- ~20,000 SMALL AREAS COVERING IRELAND
- EACH COVERS APPROX. 200 PEOPLE
- EACH CENSUS FEATURE AVAILABLE AT THIS LEVEL

THIS IS THE POWER WE WERE
LOOKING FOR!



WE REALISED THAT WE COULD DO WHAT EXPERIAN DOES

BUT ALSO:

- WE WOULD HAVE THE CODE
- WE COULD INTEGRATE IT WITH ANY DATA SCIENCE PROJECT
- WE COULD TUNE IT TO FIT OUR PARTICULAR NEEDS

NOT A TRIVIAL TASK

IT'S HARD TO MAKE SENSE OF THIS MUCH DATA:

- THERE ARE 18,488 SMALL AREA MAPS
- EACH SMALL AREA MAP IS REPRESENTED BY A ROW
- EACH ROW HAS 767 ENTRIES ONE FOR EACH FEATURE



WHAT WE HAVE IS A
REALLY BIG MATRIX

WITH 18,488 ROWS & 767 COLUMNS

$$\begin{array}{cccc|c} a_{11} & a_{12} & \cdots & a_{1n} & b_1 \\ a_{21} & a_{22} & \cdots & a_{2n} & b_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} & b_n \end{array}$$

LINEAR ALGEBRA
TO THE RESCUE

SINGULAR VALUE DECOMPOSITION

REDUCES THE SIZE OF THE PROBLEM BY DESCRIBING THE DATA IN A NEW SET OF AXES

- FEATURES OFTEN PARTIALLY CORRELATE
- IF YOU KNOW ONE COLUMN, YOU PARTIALLY KNOW THE OTHER
- SO WE DESCRIBE BOTH USING A SINGLE COLUMN (WITH SOME MINOR LOSS OF INFO)

$$M = \begin{bmatrix} M_{1,1} & M_{1,2} \\ M & M \end{bmatrix}$$



SVD LETS US SHRINK DOWN TO 100 COLUMNS

- AND RETAIN 80% OF THE FULL INFORMATION
- THIS IS STILL PRETTY HARD TO VISUALISE

A scatter plot consisting of numerous small, semi-transparent blue dots arranged in a roughly circular pattern. The dots are set against a light gray background with a white grid. The overall shape is somewhat irregular, with some gaps and a slightly jagged edge. The text 'VISUALISATION IN DATA EXPLORATION' is centered over the plot in a black, sans-serif font.

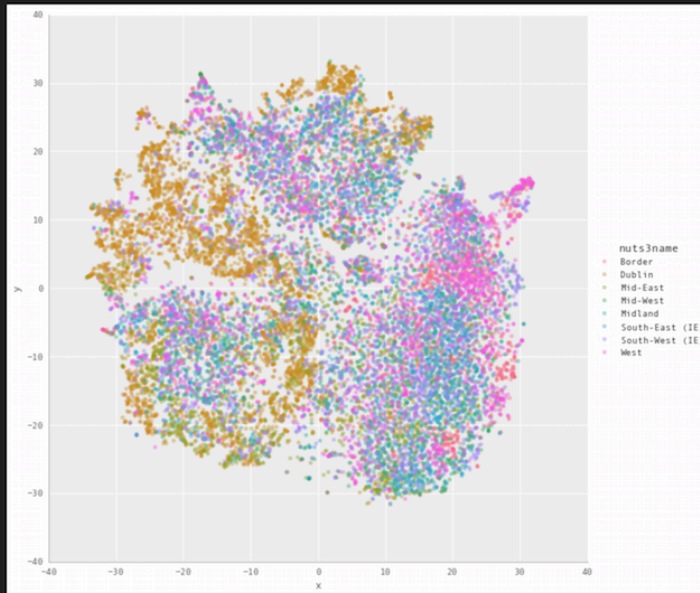
VISUALISATION IN DATA EXPLORATION

HUMAN INTERPRETATION IS OFTEN VITAL

WE WANT TO SEE STRUCTURE IN THE DATA, BUT
VIEWING 100 DIMS IS *STILL* TRICKY

- USE UNSUPERVISED LEARNING
- T-DISTRIBUTED STOCHASTIC NEIGHBOR EMBEDDING (T-SNE)
- CREATE A 2D REPRESENTATION OF ND SPACE

OVERLAY KNOWN CLASSES

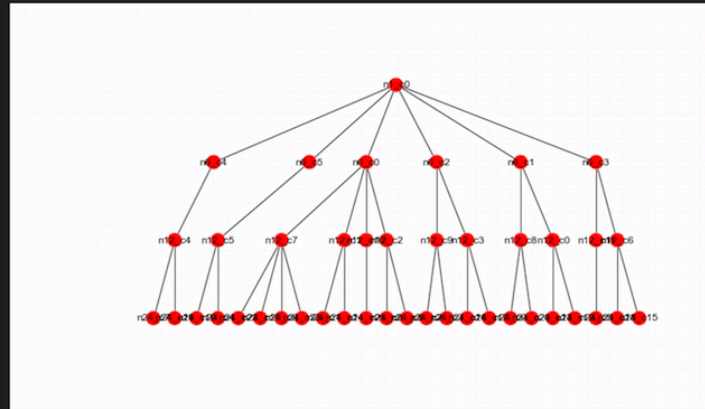


- Instantly see grouping in the regions
- The tSNE was fitted using full census data, but *not* the region id
- Yet we see the regions have similarity

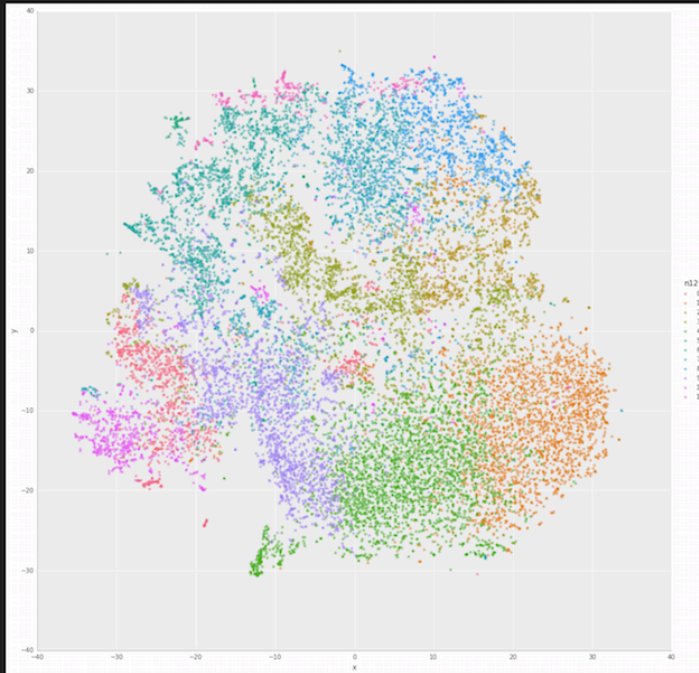
A GOOD HINT THERE'S
REAL STRUCTURE IN
THE DATA

AGGLOMERATIVE HIERARCHICAL CLUSTERING

- Group nearby datapoints into progressively larger clusters
- Get a nested hierarchy of clusters
- Choose your level



INTERESTING STRUCTURE!



- Clustering was entirely unsupervised
- i.e. determined only by the data itself
- Now we need to understand what the clusters mean

A close-up photograph of a vibrant red pushpin with a silver metal stem, firmly stuck into a colorful, detailed map. The map shows various geographical features and colors, though they are slightly blurred in the background. The pushpin is positioned on the right side of the frame, pointing towards the center.

FIRST TRICK: MAP THEM

COLOR THE SMALL AREAS BY CLUSTER



IRELAND

A vibrant, multi-colored map of Dublin, Ireland, where each area is filled with a different hue like purple, green, yellow, and red. The map is overlaid with a white grid pattern. The word "DUBLIN" is written in large, white, sans-serif capital letters across the center of the map. To the right, a dark blue area represents the sea, with a dark grey silhouette of a coastline or harbor structure extending into it.

DUBLIN



WOW



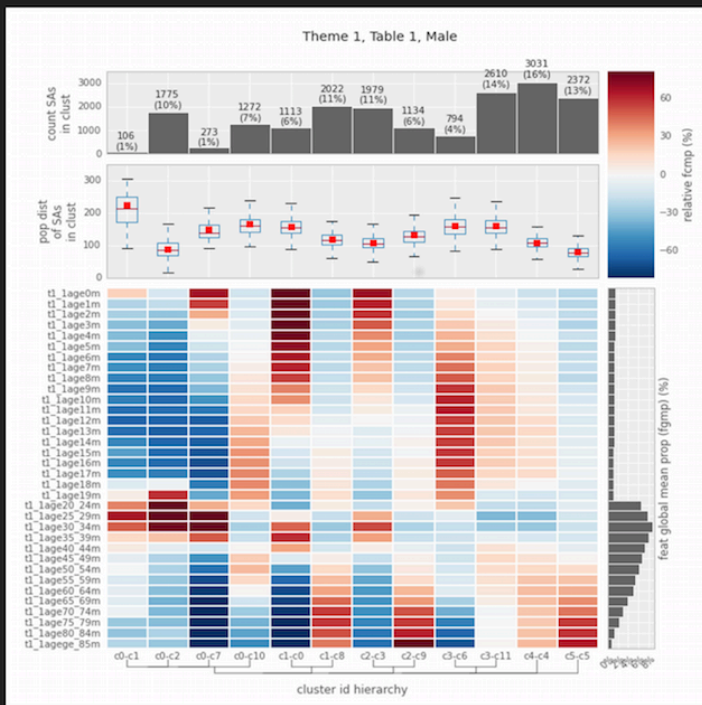
BUT WHAT DO THE CLUSTERS MEAN?

IF YOU LIVE IN IRELAND YOU CAN MAKE A GOOD
GUESS...



SECOND TRICK:
PONDERING

INTERPRETING THE CLUSTERS



- Custom visualisation of raw feature proportions per cluster
- Requires lots of careful evaluation
- Incorporate expert opinion



RESULT

VALUABLE GEO-
SOCIOECONOMIC
INDICATORS FOR USE
IN OTHER PROJECTS



PART TWO

BRINGING STATISTICAL MODELING INTO
THE BOARDROOM

-

WHITE-BOX MODELLING

SURVIVAL ANALYSIS

BAYESIAN INFERENCE



WHITE-BOX MODELLING

THE ABILITY TO INTROSPECT THE MODEL FUNCTION

"WHAT'S OUR CHURN RATE?"

SEVERAL DIFFERENT WAYS TO MEASURE CHURN
PARTICULARLY IMPORTANT IN LIFE INSURANCE

EUREKA LIFE INSURANCE COMPANY of Baltimore, Md.

INFANTILE BONUS ADDITIONAL POLICY

Table A

Amount payable if death occur after policy has been in force the following periods. Weekly Premium Ten Cent

Under six Months	Six Months	One Year	Two Years	Three Years	Four Years	Five Years	Six Years	BONUS ADDITIONS		
								Paid up and 10 years	Plus Additions	Maximum Amount
\$16	\$ 30	\$ 50	\$ 70	\$ 90	\$110	\$160		38	\$28	\$268
30	50	70	90	110	160	240		40	30	270
40	70	90	110	160	240			42	32	272
50	90	110	160	240				44	34	274
60	110	160	240					46	36	276
70	130	240						48	38	278
80	240							50	40	280
90	240							52	42	282
100	240							54	44	284

"WHAT'S OUR PROJECTED CREDIT RISK?"

A FAR MORE JUICY QUESTION

... AND IT NEEDS A DEFENSIBLE ANSWER

After policy has been in force ten years, an amount equal to ten times the weekly premium will be added to amount payable at death, and on each anniversary of policy thereafter an equal amount will be added until the maximum amount has been reached; after which, the policy becomes fully paid up and no further premiums will be required.



CONVENTIONAL METHODS

MS EXCEL

RUNNING AVERAGES & POINT-ESTIMATES

VERY LIMITED INTROSPECTION, EXPLANATION AND
FORECASTING



TIME-TO-EVENT MODELING

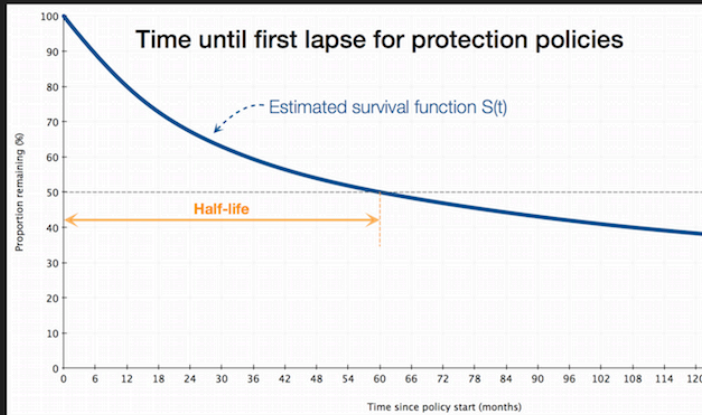
"HOW MUCH TIME ELAPSED BETWEEN EVENTS?"

- HALF-LIFE OF A RADIOACTIVE COMPOUND
- DURATION OF A PITSTOP IN F1
- DURATION BETWEEN A CUSTOMER JOINING AND QUITTING A CONTRACT

SURVIVAL ANALYSIS

A WELL-STUDIED PART OF MEDICAL STATISTICS
BUT NOT USUALLY DISCUSSED IN MACHINE LEARNING
SEEMS TO HAVE BEEN OVERLOOKED

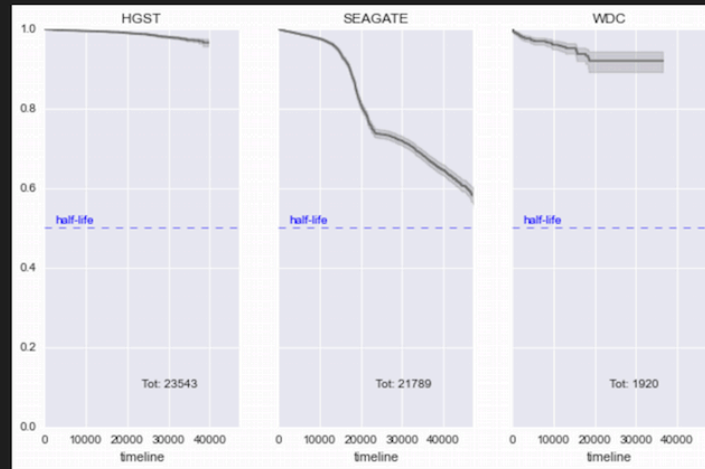
THE BASIC IDEA:



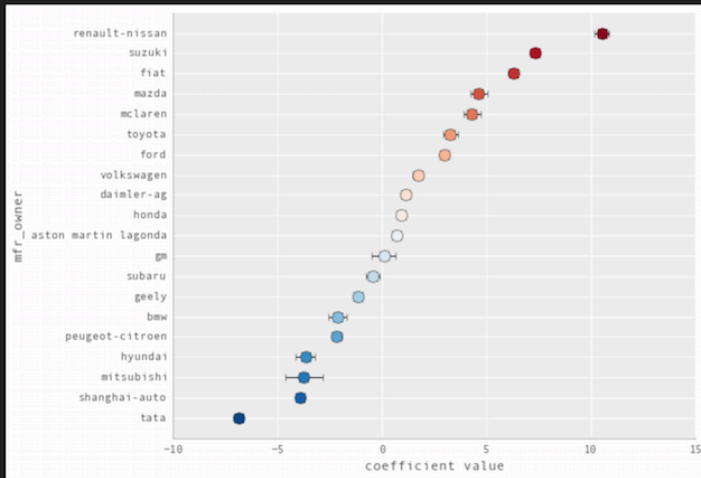
- Take a cohort at some start date
- Count the prop. remaining at subsequent dates
- Simply draw or fit the line

SIMPLEST VERSION: KAPLAN-MEIER

- Simple count-based description of events
- Draw the line per group
- Prediction not possible



A MODEL-BASED VERSION: COX PROPORTIONAL HAZARDS



- Fit semi-parametric linear model
- Learn effect of feature values
- Make predictions on new data



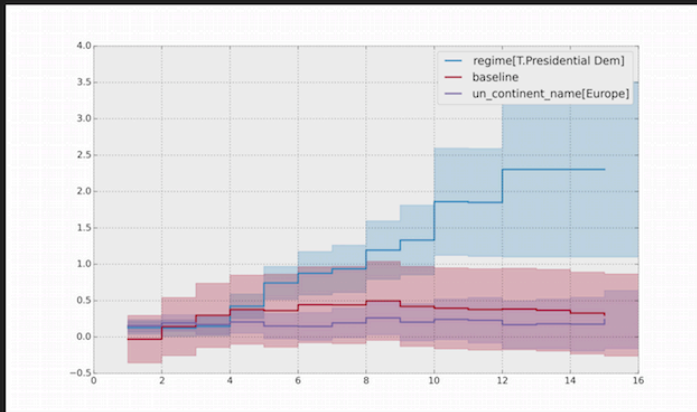
THIS IS ALREADY SUFFICIENT TO ANSWER

WHAT IS THE EXPECTED RANGE OF SURVIVAL OVER
TIME?

CAN I PREDICT THE SURVIVAL RATE?

WHAT ARE THE DRIVERS OF THE SURVIVAL RATE?

INCREASE THE COMPLEXITY TO BETTER FIT THE REAL-WORLD



- Survival Regression
- Allow hazards to vary with time: Aalen Additive
- Bayesian approach: Gaussian Processes

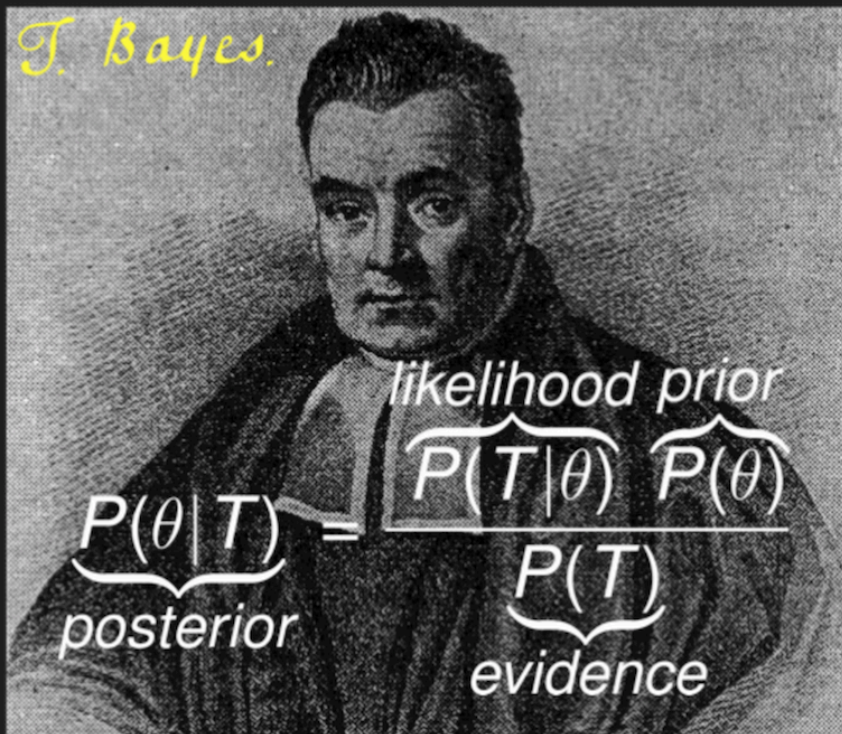
The image features two dice on a dark, textured background. The die on the left is purple and shows a face with three white dots. The die on the right is gold and shows a face with five white dots. The text is centered over the dice.

THE BEST MODELS
PRESERVE
UNCERTAINTY

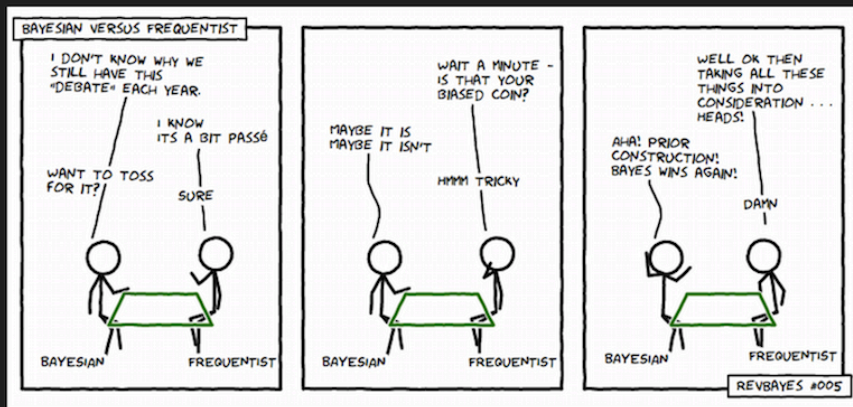
BAYESIAN INFERENCE

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

T. Bayes.

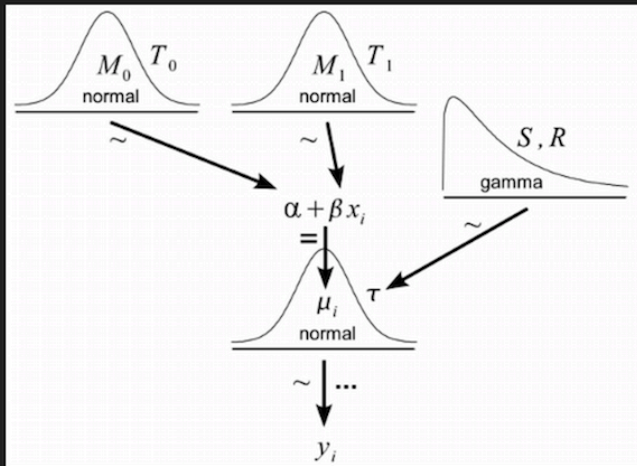
A black and white portrait of Thomas Bayes, a man with a high-collared coat and a serious expression. The portrait is the background for the text and mathematical symbols.
$$\underbrace{P(\theta|T)}_{\text{posterior}} = \frac{\overbrace{P(T|\theta)}^{\text{likelihood}} \overbrace{P(\theta)}^{\text{prior}}}{\underbrace{P(T)}_{\text{evidence}}}$$

A NATURAL ALTERNATIVE TO TRADITIONAL (FREQUENTIST) STATISTICS

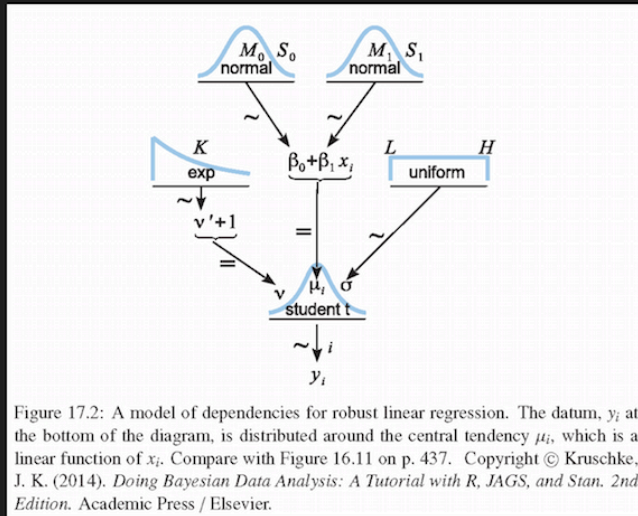


MADE POSSIBLE THROUGH COMPUTATIONAL POWER

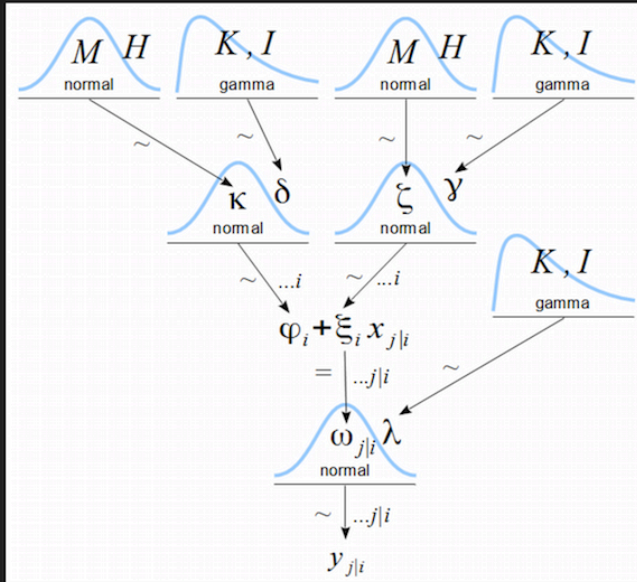
FAR TOO MUCH DETAIL FOR HERE



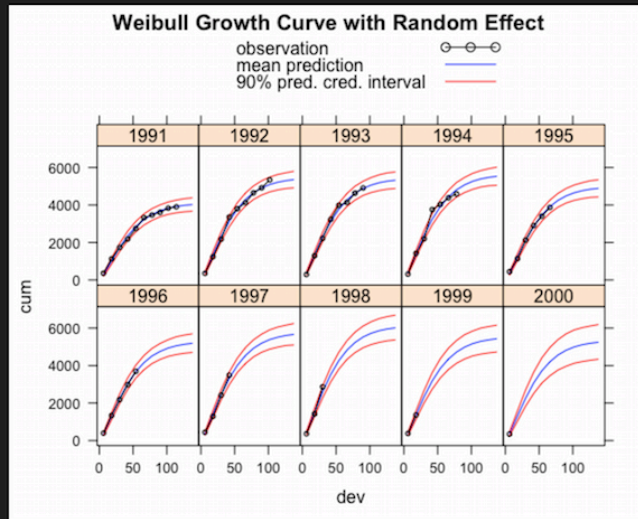
FAR TOO MUCH DETAIL FOR HERE



FAR TOO MUCH DETAIL FOR HERE

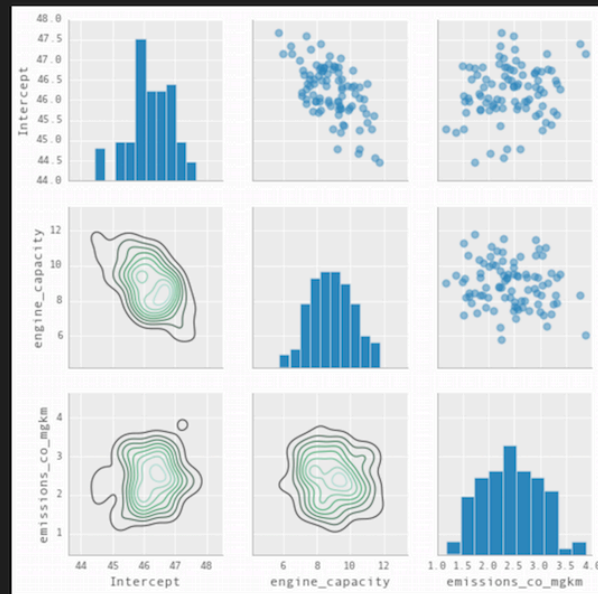


FAR TOO MUCH DETAIL FOR HERE

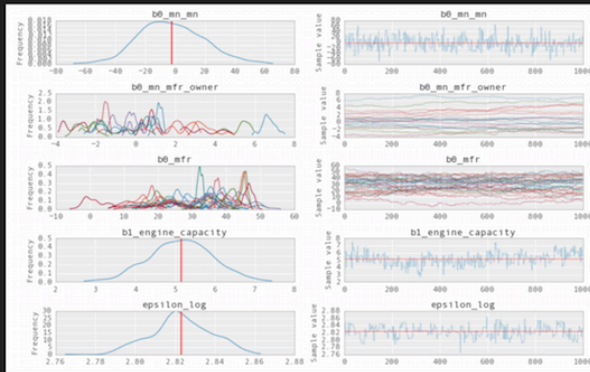


BIGGEST WIN: MODEL INTROSPECTION

- Design structure of model to reflect real processes & relationships
- Inform the model using prior knowledge
- Use qualified results and predictions
- Maintain uncertainty



STATE OF THE ART ACADEMIC TOOLS & TECHNIQUES OPEN TO ALL



- Basic non-sampling frameworks: scikit-learn, arm, nlme etc
- Advanced model frameworks and samplers: STAN, PyMC3, emcee
- Probabilistic graphical models: gRain, pgmpy



THIS TRULY DOES OFFER

EXTRAORDINARY
POWER



ALLOWING US TO CREATE MODELS THAT:

ACCURATELY REFLECT REAL-WORLD PROCESSES

PRESERVE UNCERTAINTY IN PREDICTION

WORK WITH SMALL AND OBSCURE DATASETS

AUTOMATICALLY IMPROVE WITH MORE DATA



ALLOWING US TO ANSWER
QUESTIONS LIKE:

KNOWING THE YEARLY CUMULATIVE POLICY CLAIMS
IN MY INSURANCE BOOK OVER THE PAST 10 YEARS,
HOW MUCH SHOULD I SET ASIDE FOR CLAIMS
RESERVES FOR THIS YEAR'S NEW COHORT, AND TO
WHAT RANGE EITHER SIDE?



ALLOWING US TO ANSWER QUESTIONS LIKE:

BASED ON PRIOR APPLICATIONS FOR LIFE INSURANCE POLICIES, CAN I DETERMINE WHICH PROSPECTIVE CUSTOMERS TO SEND FOR MEDICAL EXAMINATIONS (EXPENSIVE, NOT OFTEN NECESSARY), DEFINE BETTER QUESTIONS TO ASK IN THE PRE-SCREENING, AND CHANGE MY BELIEFS BASED ON NEW EVIDENCE?

A man in a hooded tunic with a sun emblem on the chest, carrying a large bundle on his back, is the central figure. He is looking down and to the left. The background shows a medieval setting with a stone building and other figures. The text "BUT REMEMBER:" is overlaid in white, sans-serif font.

BUT REMEMBER:

IT'S ONLY A MODEL



PART THREE

APPLYING MORE EXOTIC MACHINE
LEARNING ACROSS THE BUSINESS

-

NATURAL LANGUAGE PROCESSING
UNSUPERVISED FEATURE REDUCTION
PATTERN MATCHING AND PREDICTION



"MACHINE LEARNING" SOUNDS

OMINOUS

VAGUE

MONEYPIT

"OUR COMPANY HAS BEEN USING
DATA FOR YEARS"

~~SQL & SPREADSHEETS~~

~~BI DASHBOARDS~~

~~EXPERT / RULE BASED SYSTEMS~~

A yellow and black toy robot with large camera eyes is positioned on the left side of the image, sitting on a computer keyboard. The robot has a yellow body with black accents and large, round camera lenses for eyes. The keyboard is a standard grey and white layout, and the entire scene is set against a dark, slightly blurred background.

WE'RE TALKING ABOUT INTELLIGENTLY LEARNING FROM DATA

TRAINING MODELS TO PREDICT OPERATIONAL
OUTCOMES

DISCOVERING CORRELATIONS, MATCHING PATTERNS
OF BEHAVIOR

RUNNING THE BUSINESS MORE EFFECTIVELY

A yellow and black toy robot with large camera eyes is positioned on the left side of the image, resting on a computer keyboard. The robot has a boxy body and a small antenna. The keyboard is a standard grey and white design. The background is a dark, slightly blurred grey.

WE'RE TALKING ABOUT INTELLIGENTLY LEARNING FROM DATA

SELLING INSIGHTS AND DATA PRODUCTS BACK INTO
THE MARKET

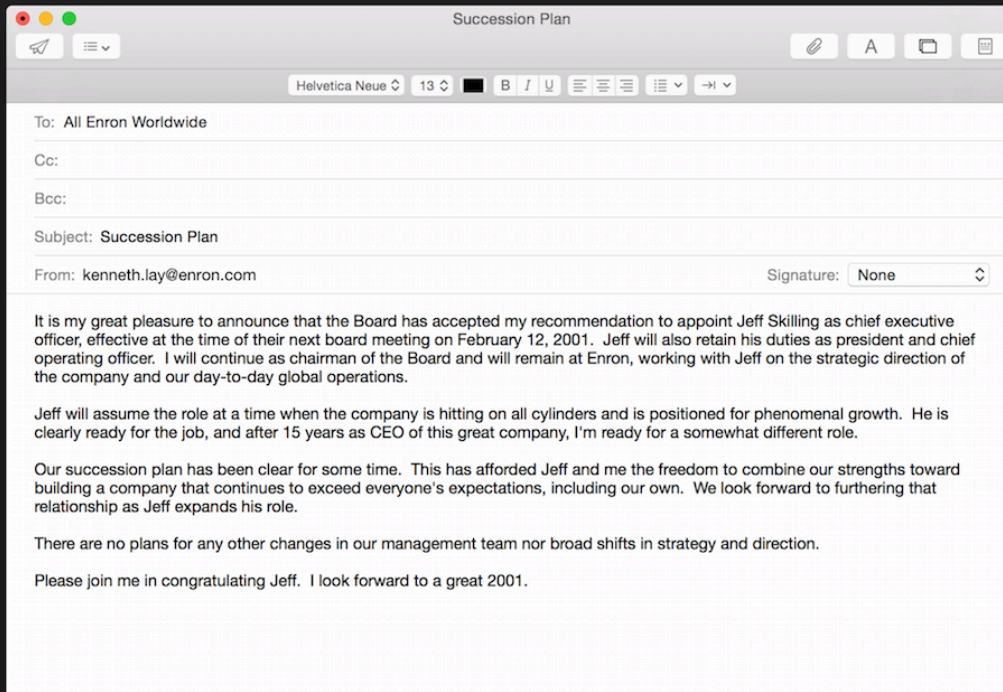
EXTRACTING INFORMATION FROM ALL THAT BIG DATA
YOU'RE COLLECTING

A top-down view of a cluttered desk. In the foreground, there's a box of Amoxicillin 500 mg capsules, a silver can, a blue highlighter, and a pair of black scissors. Several papers and documents are scattered around, some with text and tables. In the background, two large open books are visible, with hands pointing at pages. The overall scene suggests a busy workspace or a place of intense research and data analysis.

THERE'S A HUGE AMOUNT OF
DATA ALREADY INSIDE THE
BUSINESS

IT JUST TAKES A LITTLE PUTTING TOGETHER

NLP UPON UNCONVENTIONAL DATA SOURCES



AFTER CAREFUL CLEANING AND PREPROCESSING

Office Memorandum • UNITED STATES GOVERNMENT

TO : Mr. Fletcher *W.F.*

DATE: July 1, 1949


FROM : V. P. Keay *VPK*

SUBJECT:

Mr. Tolson _____
Mr. Clegg _____
Mr. Glavin _____
Mr. Ladd _____
Mr. Nichols _____
Mr. Rosen _____
Mr. Tracy _____
Mr. Egan _____
Mr. Gurnea _____
Mr. Harbo _____
Mr. Mohr _____
Mr. Pennington _____
Mr. Quinn Tamm _____
Tele. Room _____
Mr. Nease _____
Miss Gandy _____

Referral
① William L. Uanna, who was formerly Chief of Personnel Security, Atomic Energy Commission, but who is presently connected with the U. S. Engineers in Kansas City, called me from Kansas City on June 30, 1949.

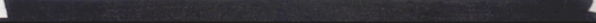
82
1



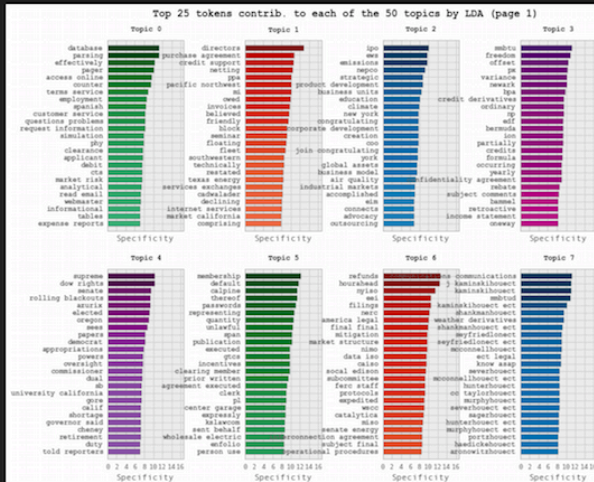
I told Uanna we would do the best we could but, of course, we could not promise any special time on these investigations.

RECOMMENDATION:

Referral
It is suggested this memorandum be referred to Mr. *Chian* in Division Six in order that special attention can be given to these requests when they arrive.



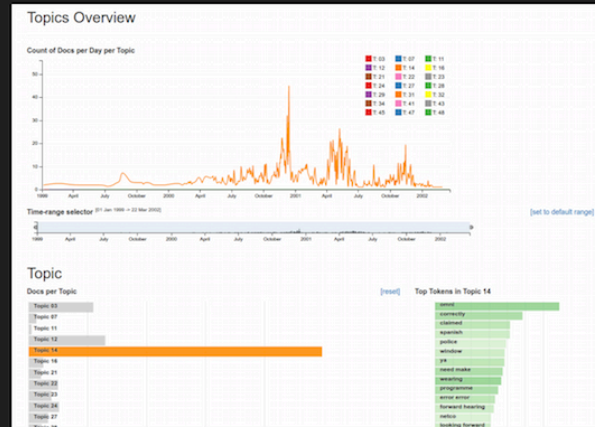
MODEL THE TOPICS OF CONVERSATION WITHIN A CORPUS



- Topic modelling (Latent Dirichlet Allocation)
- Cluster words into topics by their co-occurrence
- Gain a new way to define the contents of documents

USE TOPICS AS A METHOD TO INTROSPECT THE HEARTBEAT OF COMMUNICATIONS

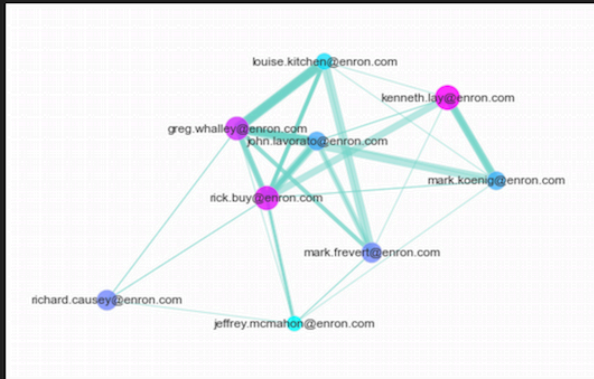
- View trends of topics by time
- Correlate topics to internal / external events
- Use topics in other models e.g. predict customer satisfaction & longevity according to topics





CAN WE USE SOCIAL
& ORGANISATIONAL
STRUCTURE TOO?

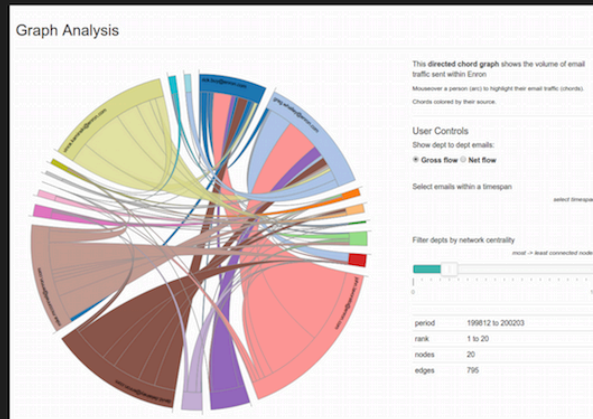
USE THE FROM: AND TO: TO CREATE A NETWORK / GRAPH



- Nodes are email accounts / people
- Edges are the communications between them
- Overall counts / frequency / time of day / season

ATTACH THOSE TOPICS FROM BEFORE TO GAIN A RICH PICTURE OF COMMUNICATIONS

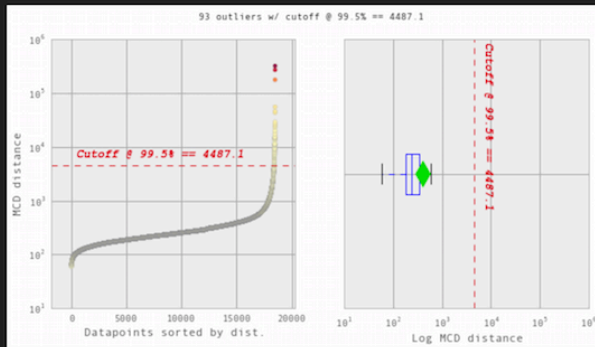
- Non-invasive introspection of the state of the organisation
- Relevant to networks of brokers, employees, customers, reinsurers etc.



A cartoon illustration of two Whodunnits from the Dr. Seuss book 'The Cat in the Hat's Learning Library'. They are depicted as a pair of identical, blue, spiky-haired creatures with large noses and wide, happy smiles. They are wearing matching brown, long-sleeved jumpsuits with a white circular patch on the chest that contains the word 'WHODUNNIT'. The characters are shown in a dynamic, running pose, with their arms and legs outstretched. The background is a plain, light gray color, and the ground is indicated by several horizontal black lines at the bottom of the frame.

A FEW MORE THINGS
NOT YET MENTIONED

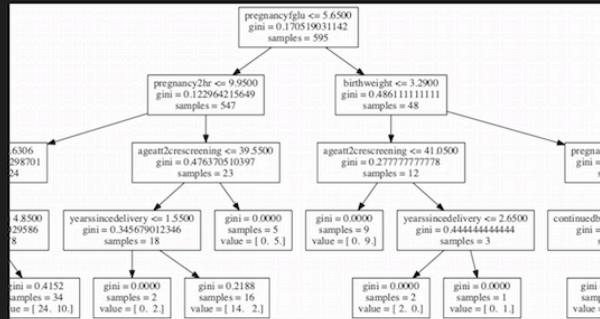
OUTLIER / FRAUD DETECTION



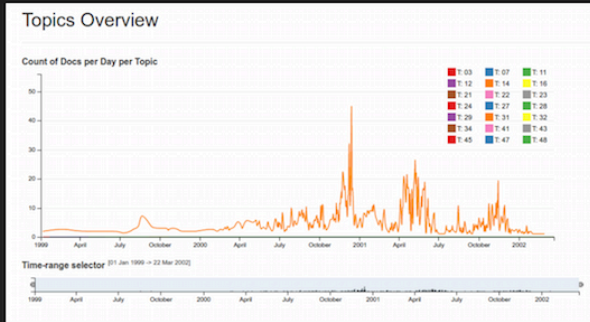
- Describing data in a vector space and assigning outlier flags to unusual points
- (or use graphs / trees)
- Triage these first in fraud or operations investigations

SUPERVISED CLASSIFICATION

- Good old-fashioned prediction
- Product-market fit
- Operational improvements
- Meeting compliance etc.



INTERACTIVE DASHBOARDING



- Surface descriptive & predictive insights to the business to enable better decision making
- Lightweight Javascript frameworks are powerful and easy to use



WE'VE ONLY
SCRATCHED THE
SURFACE TODAY

BUT MANY COMPANIES ARE ALREADY DOING THIS

IN SUMMARY

DATA SCIENCE IS APPLICABLE
THROUGHOUT THE INSURANCE BUSINESS

WE ARE PROVING THIS EVERY DAY THROUGH OUR
PROJECTS

1. CURATING EXTERNAL DATA TO BETTER UNDERSTAND CUSTOMERS
2. BRINGING STATISTICAL MODELING INTO THE BOARDROOM
3. APPLYING MORE EXOTIC MACHINE LEARNING ACROSS THE BUSINESS

A wooden shelf holds a collection of tools, including screwdrivers with various handles (wood, black, red) and other hand tools. A white label with the numbers '173456789' and a red logo is attached to the shelf. The background is a plain wall.

THE TOOLS ARE POWERFUL AND
LARGELY FREE

A blurred background image of a workshop. In the foreground, a wooden shelf holds several screwdrivers with different handle colors (black, wood, red). Below the shelf, a white label with the numbers '173456789' and a red square icon is visible. The background shows more tools and a wooden cabinet, all out of focus.

THERE'S A WEALTH OF
INFORMATION ALREADY INSIDE
THE BUSINESS



AND WITH INTELLIGENT
CURATION, HIGH QUALITY
EXTERNAL DATA IS READILY
AVAILABLE



SPEND MONEY ON THE TEAM



ENABLE THEM TO HONE & SHARE
SKILLS & EXPERTISE



PLOUGH INSIGHTS BACK INTO THE
BUSINESS

Ech. 02 020 P.M.

Fig. 349. — *Batteuse Damey à manège direct placé sous la batteuse.*



AND INCORPORATE DATA
SCIENCE THROUGHOUT

Ech. 02 020 P.M.

Fig. 349. — Batteuse Damey à manège direct placé sous la batteuse.

APPLIED

AI

