Advanced Analytics

Predictive Analytics and AI:
the future of actuaries?

Antoine LY
antoine.ly@milliman.com

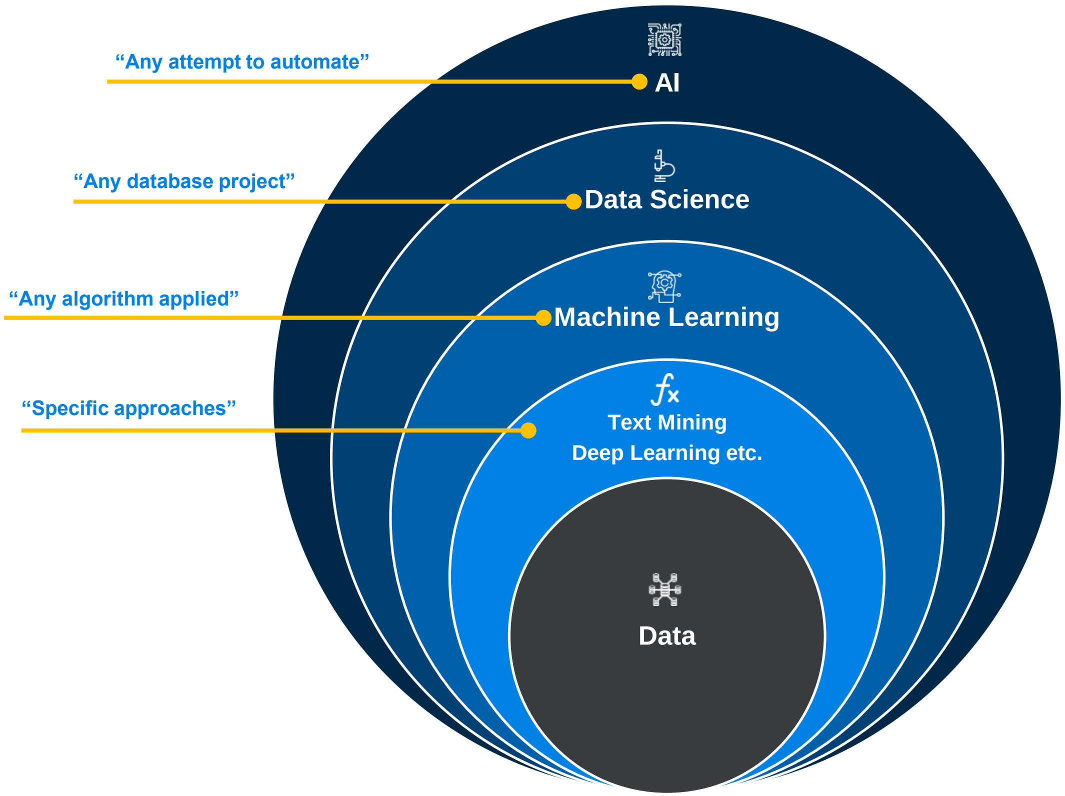
December 2018 – Tel Aviv

Milliman

Advanced Analytics

Artificial Intelligence?

- Origin: Alan Turing (1950).
- Today, **many definitions**:
  - Big Data + Machine Learning = AI,
  - AI = Handle data with intelligence,
  - Etc.
- This covers **algorithms**, **automation**, **data visualization**,…
- Recent **buzz words**: deep learning, GDPR, discriminative algorithm, …



Advanced Analytics

General presentation

- Advanced analytics combines **multiple fields of expertise** (Predictive modelling, IA, IT) to tackle complex issues which involve diverse sources of data and strong structural constraints
- The diversity of profiles and experience of the Milliman Analytics team allows us to work on **different topics** and ensures the success of projects.

Predictive modelling

Projects with **machine learning** and **actuarial** modelling

Process automation

Design of tools and dashboards in R (**Shiny**), Python (**Dash**, Flask)

Data management

Management of different data sources, **big data** processing, use of structured and unstructured data

Software development

Development of **SaaS softwares**, use of best practices (CI, unit tests, git), local or cloud deployment

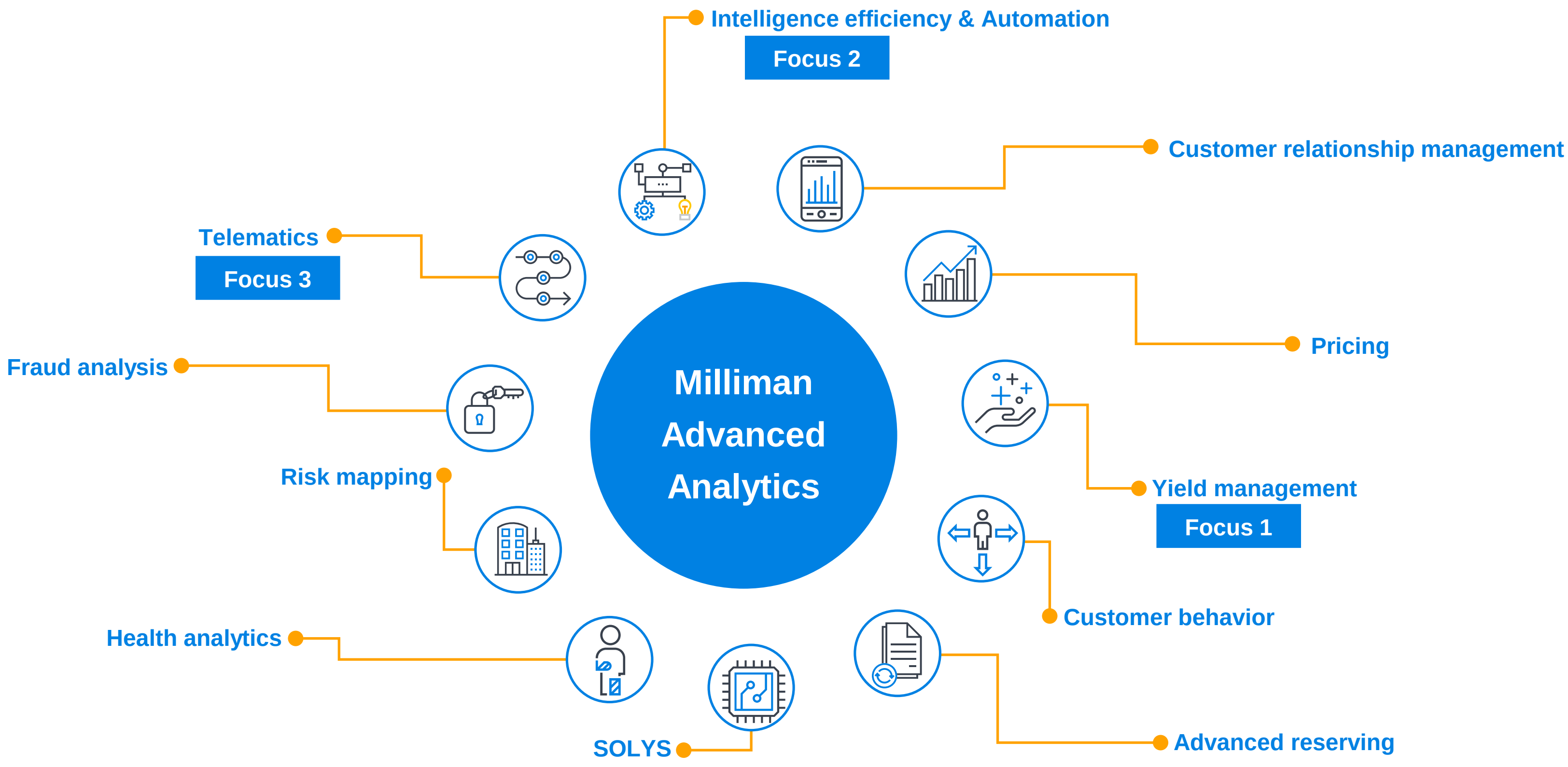
Advanced Analytics

Our vision: Key points

- Milliman has been working on different analytics projects for 5 years. From our experience ranging from studies to tool developments, we believe that successful analytics in a large company depends on these **key points**:

Advanced Analytics

Different areas



Focus 1
Yield Management

Artificial efficiency applied to yield management

Traffic and Budget forecasting for non-insurance company (1/2)

Description:

A European transportation company wants to deploy an open source solution to lead their budget and predict traffic .

Constraints:

- Use only open source technologies;
- Testing and integrating a machine learning model;
- Developing a Graphical User Interface;
- Respecting current IT infra.

Artificial efficiency applied to yield management

Traffic and Budget forecasting for non-insurance company (2/2)

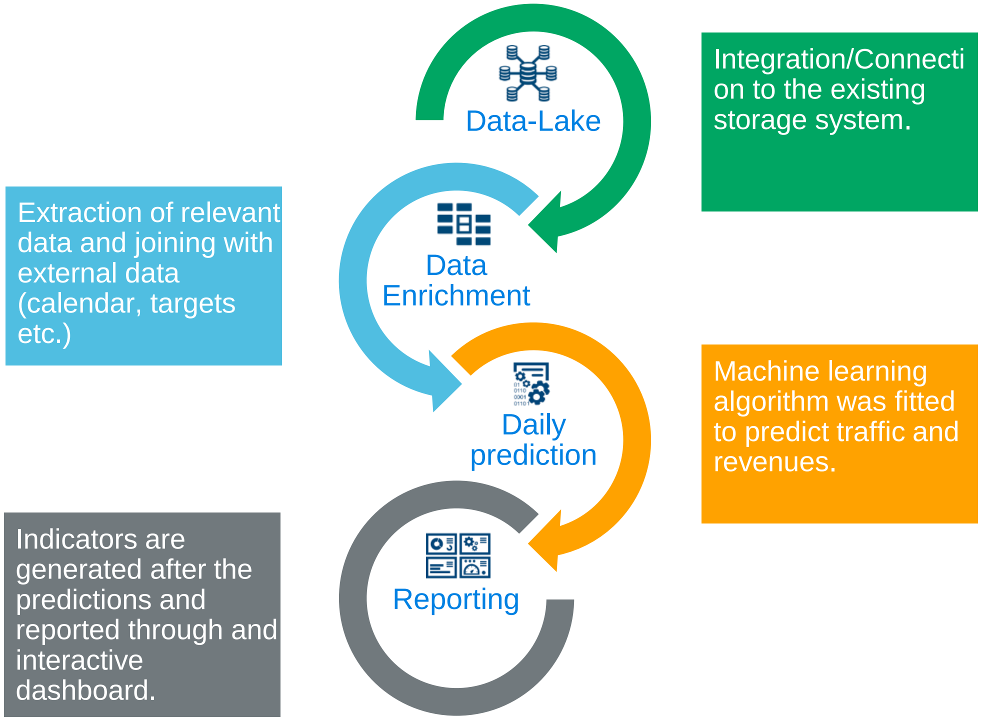
Highlights

We demonstrated that Milliman could successfully achieve a data science project. All the expected skills of a data scientist were gathered into one project: data management, machine learning modelling, web reporting, IT knowledge.

Duration of the project

2month + 6 month

(PoC + full project)



Focus 2
Text Mining

Comment analysis

Comment analysis on insurance companies

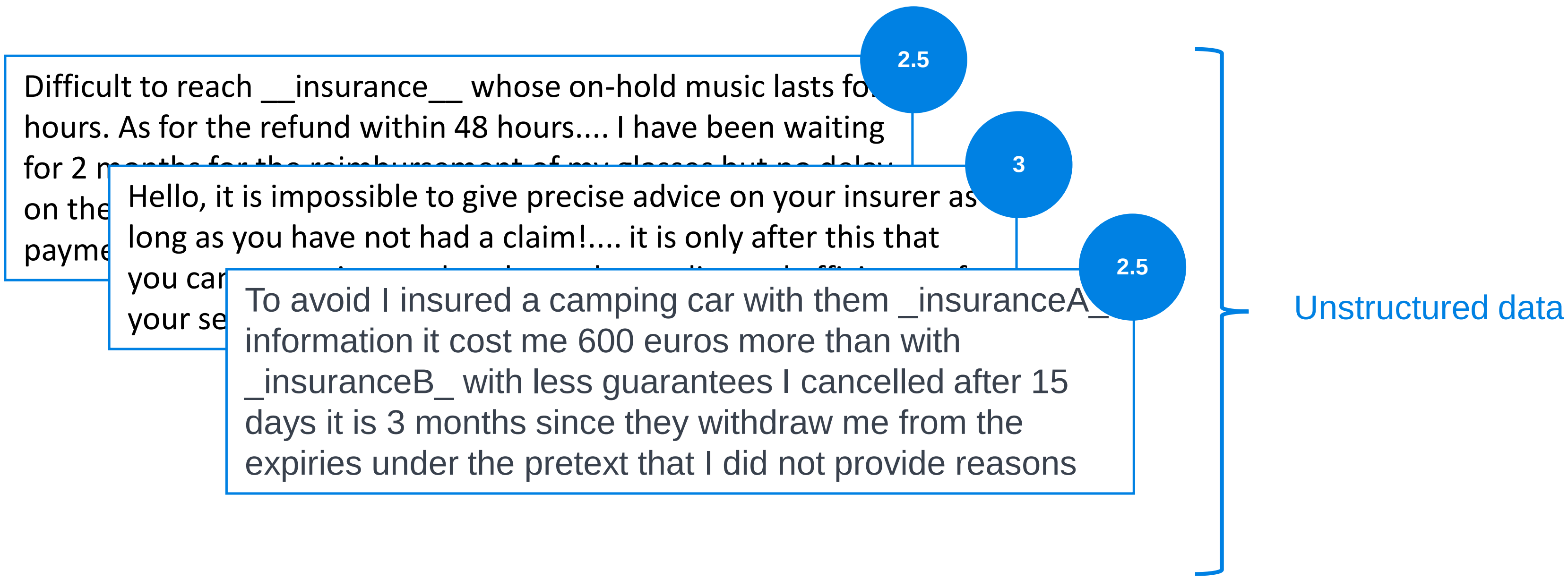
Context (1/3)

- Why analyzing comments could be useful?
  - **Clients share experience**: comments and ratings influence underwriting because people like getting feedback.
  - **Understanding user experience**: unsatisfied clients make more noise… understanding main reasons why someone can move to another insurance company is valuable.
  - **Competitive market**: new digital actors with digital services.
- How do we proceed?
  - **Lots of information**: thousands of comments are published, reading them one by one could be time consuming and not efficient to extract main topics.
  - **Automating analysis and reporting**: the programming languages used in machine learning provide a large amount of tools to ease exploration.

Comment analysis on insurance companies

Context (2/3)

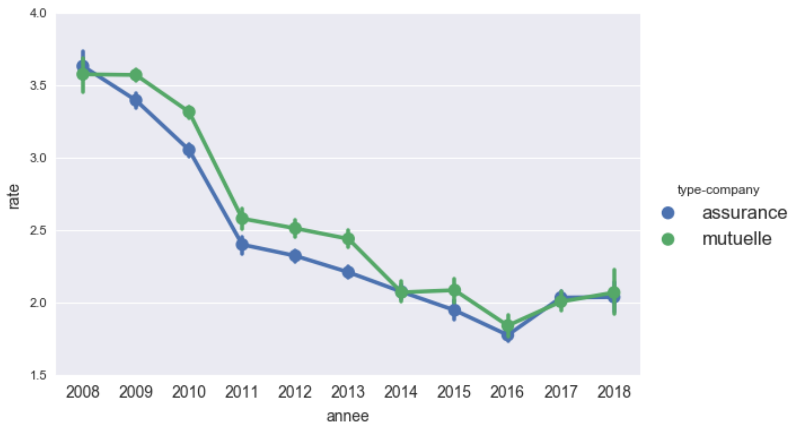
- A study lead on almost 40k comments about a mix of 180+ French insurance companies
- Each comment is rated with a mark of five points

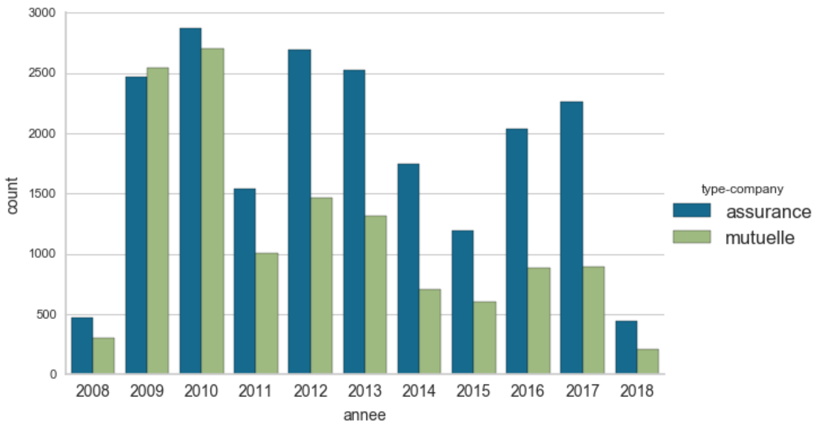


Comment analysis on insurance companies

Context (3/3)

- General trend: rate decreases 1.5 points lost in average in 10 years
- No specific distinction between mutual or insurance companies





-
- Can we learn more?

Comment analysis on insurance companies

Data preprocessing (1/2)

How to deal with text information?

no refunds from September to March despite multiple registered letters left unanswered. the people with whom we are connected by phone are never able to answer requests due to computer problems, holidays, illness... NO customer services! more than dubious practice! do not subscribe!!!!!!

[no , refund , september , multiple , letter , left , despite , unanswer , people , whom , we , be , connect , phone , never , able , request , due , computer , problem , holidays , demand , illness , service , more , dubious , practice , not , subscribe]

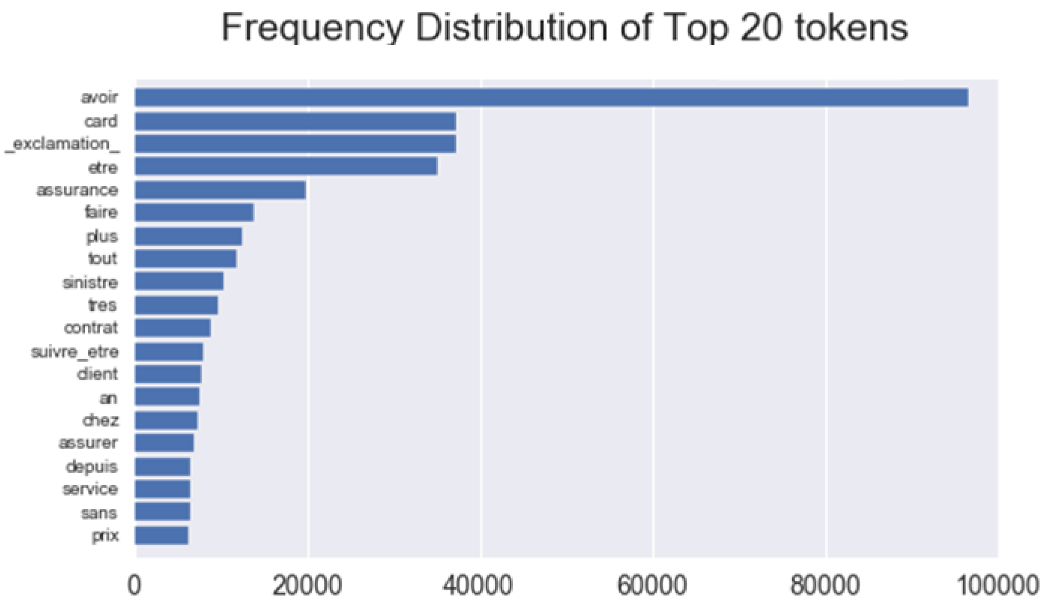
- Convert string sequence into list of words
- Stemming
- Remove “stopwords”
- Remove special characters
- Apply custom scrubbing

Comment analysis on insurance companies

Data preprocessing (2/2)

First analysis

- People who comment are **very expressive**: 1 exclamation point in average by comment
- People are **very descriptive**, large a amount of qualifying verbs and numbers to describe prices and times
- Different perspectives for the analysis:
  - Focus on **figures** (time or prices)
  - Define a **normed rating** (retreated from exposure, etc.)
  - Extract **topics** from text
  - etc.



Comment analysis on insurance companies

Text representation

Methodologies overview

[aucun , remboursement , septembre , a , mars , malgres , multiple , courrier , recommande , laisser , sans , reponse , personne , sommer_etre , mettre , relation , telephone , etre , jamais , capable , repondre , demande , suite , avoir , problem , informatique , conge , maladie , aucun , service , client , pratiquer , plus , douteur_douteux , avoir , surtout , souscrire]

Frequency

Bag of Words

dictionary aucun, avoir, aller, …, jamais, …, problem, …

Sentence	1	2	0	…	1	…	1	…

Each sentence is represented by a vector of length 37896 counting the occurrence of each word

Context

Vector representation: « Word2Vec »

dictionary aucun, avoir, aller, …, jamais, …, problem, …

word	0	0.01	2^e-4	…	w_k	…	…	…

Each word is represented by a vector but there, each value indicates a score (probability like) that a word of the dictionary co-appears with the word of interest

Comment analysis on insurance companies

Focus on topic detection (1/2)

- Unsupervised methodology to detect topic

- Ratings are biased and subjected to personal appreciation
- Extracting topics helps in understanding reason of (dis)satisfaction
- It provides more insights to enhance client experience and improve process

➔ Example with **Latent Dirichlet Allocation:** Bayesian model to cluster words/sentences

Topic 0	Topic 1	Topic 2
tres bon	avoir ete	_exclamation_
rapport a	avoir avoir	qualite prix
bon assurance	card moi_mois	a fuir
service client	suite a	rapport qualite
card card	assurance a	fuir _exclamation_
tres cher	card jour	tout aller
bon assureur	service client	_exclamation_ etre
moins cher	card card	aller bien
etre tres	a eviter	card moi_mois
concurrence	apres avoir	a eviter
aucun suivi	avoir dire	rien _exclamation_
tres mauvais	a recevoir	client _exclamation_
a deconseiller	avoir sinistre	tres mauvais
suivi dossier	avoir envoye	_exclamation_ plus
card juillet	a jamais	_exclamation_ alors

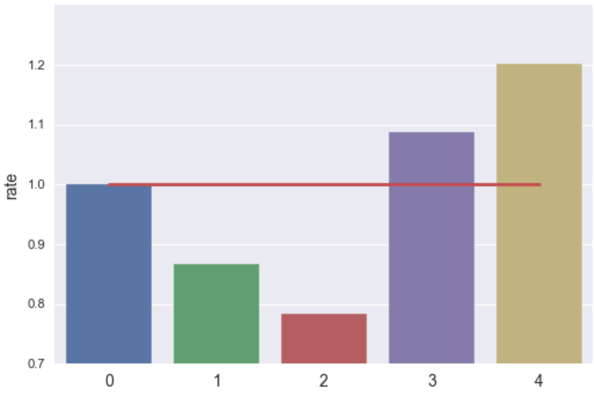
Topic 3	Topic 4
card an	cas sinistre
avoir avoir	service client
depuis card	avoir ecoute
card euro	etre tres
avoir card	prise charge
non responsable	conseiller etre
avoir etre	qualite service
avoir sinistre	prix etre
jamais avoir	tres rapide
tout risque	intervention cas
avoir jamais	assurance tres
card avoir	conseiller clientele
an avoir	disponibilite conseiller
avoir accident	qualite intervention
moins cher	

Comment analysis on insurance companies

Focus on topic detection (2/2)

Topic 0 →	Too expensive related to the provided service	e.g. «No late refund to flee (beware of accidents abroad you would still be 100% responsible) price too expensive (1200 \x80 for a Renault koelos in any risk) so the legal protection I don't even tell you »
Topic 1 →	Bad experience with one specific claim / lots of details	e.g. « […] in May 2016, I had water infiltration in two places. You are told that you must cover and that you will be reimbursed 250 euros per tarp, in fact you are not reimbursed anything[...]. You receive a letter indicating that you have had two claims and that therefore you are being charged a penalty, I only had one claim and I was not responsible for it. We tell you that in the end we only give you one claim, but we leave you the penalty We send you an estimate of the damage done at random, well below what it costs me […] »
Topic 2 →	Just a grumpy cat	e.g. « CATASTROPHIC!!!!!!!!! »
Topic 3 →	Bad experience of services	e.g. « At the time of subscription, the contracts seem to be good value for money, but this does not last. After one year of insurance without any claims, my premium is already 11% more expensive than for a new contract (with exactly the same coverage). My advisor, who was not lacking in reactivity and arguments before my subscription, remains mysteriously silent on this subject despite numerous reminders from me. In short, loyalty is not rewarded r»
Topic 4 →	Happy user	e.g. « Although a little more expensive than many of the competitors, [...] has shown excellent coverage in the event of an automobile accident requiring significant breakdown assistance and compensation thanks to globally attentive advisers and great responsiveness. »

Lift chart

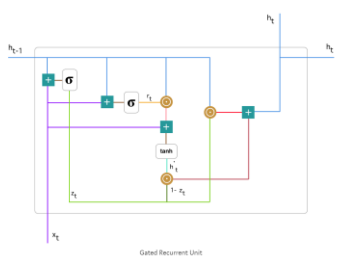


Use case: Net promoter score automation

An example to apply text-mining

Understand and design models

- **Understanding models**
- **Programming** skills
- Business knowledge



API

- Capitalizing on tool development experience
- Risk modeling
- Optimizing production process with **open source**

Dashboarding Rshiny / Dash

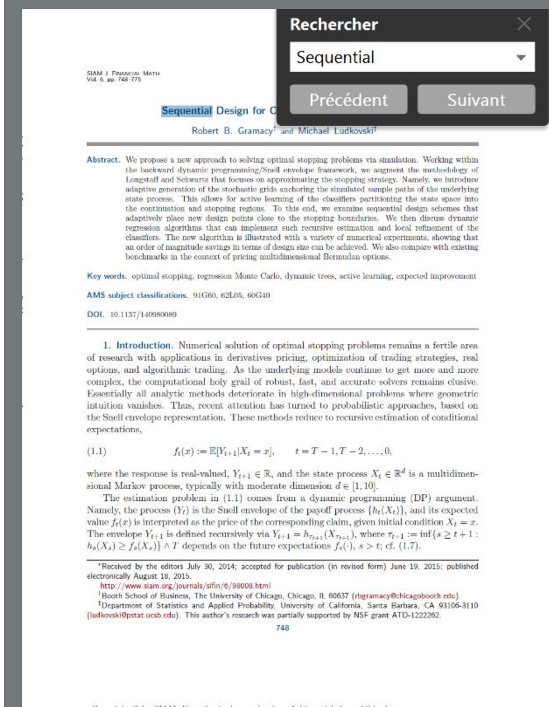
- **Monitoring results**
- **Optimize long term ROI**

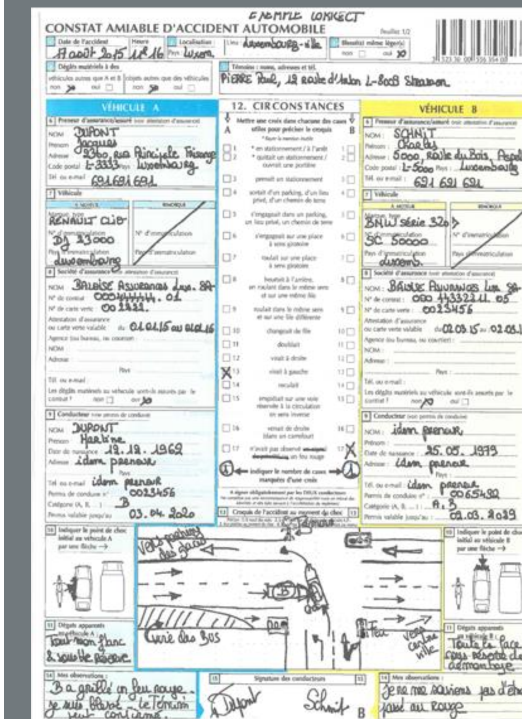
Optical Character Recognition (OCR)

OCR

Some context

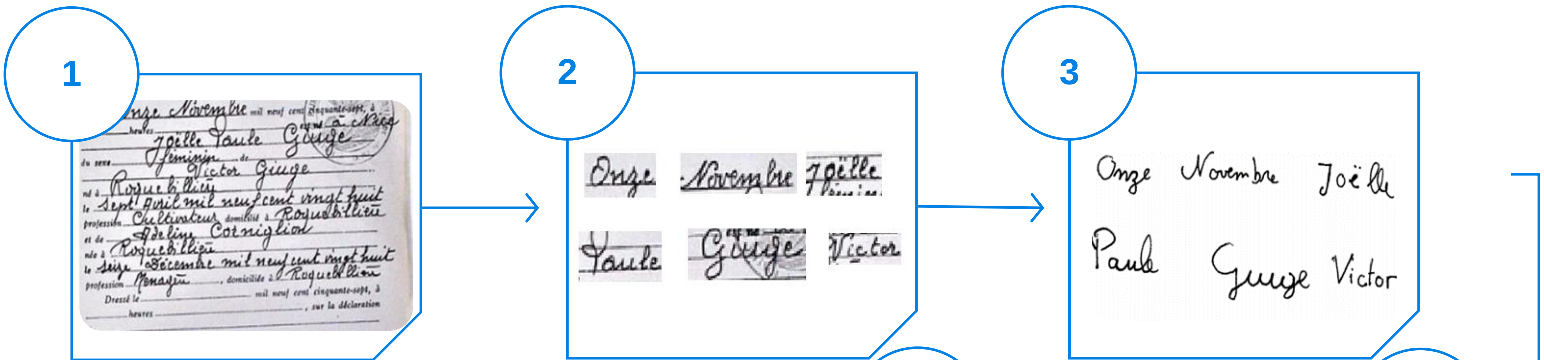
Pdf / word document	Scanned	Scanned + Handwritings

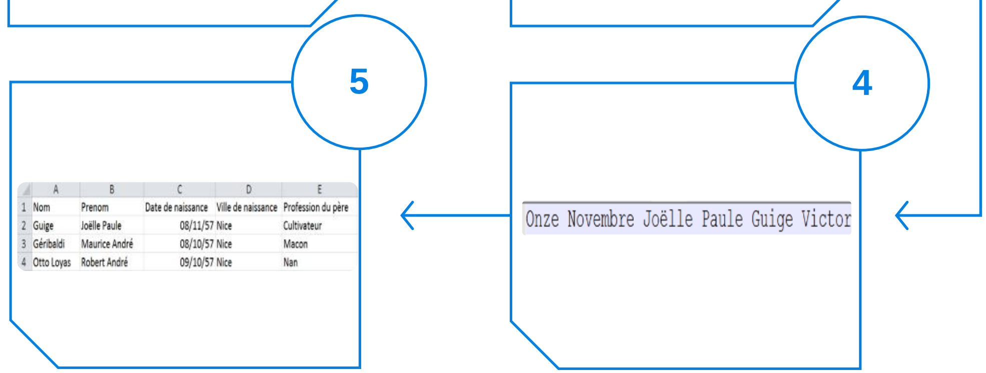




OCR algorithms with handwritten writing style

Text-Mining and document analysis





	A	B	C	D	E
1	Nom	Prenom	Date de naissance	Ville de naissance	Profession du père
2	Guige	Joëlle Paule	08/11/57	Nice	Cultivateur
3	Géribaldi	Maurice André	08/10/57	Nice	Macon
4	Otto Loyas	Robert André	09/10/57	Nice	Nan

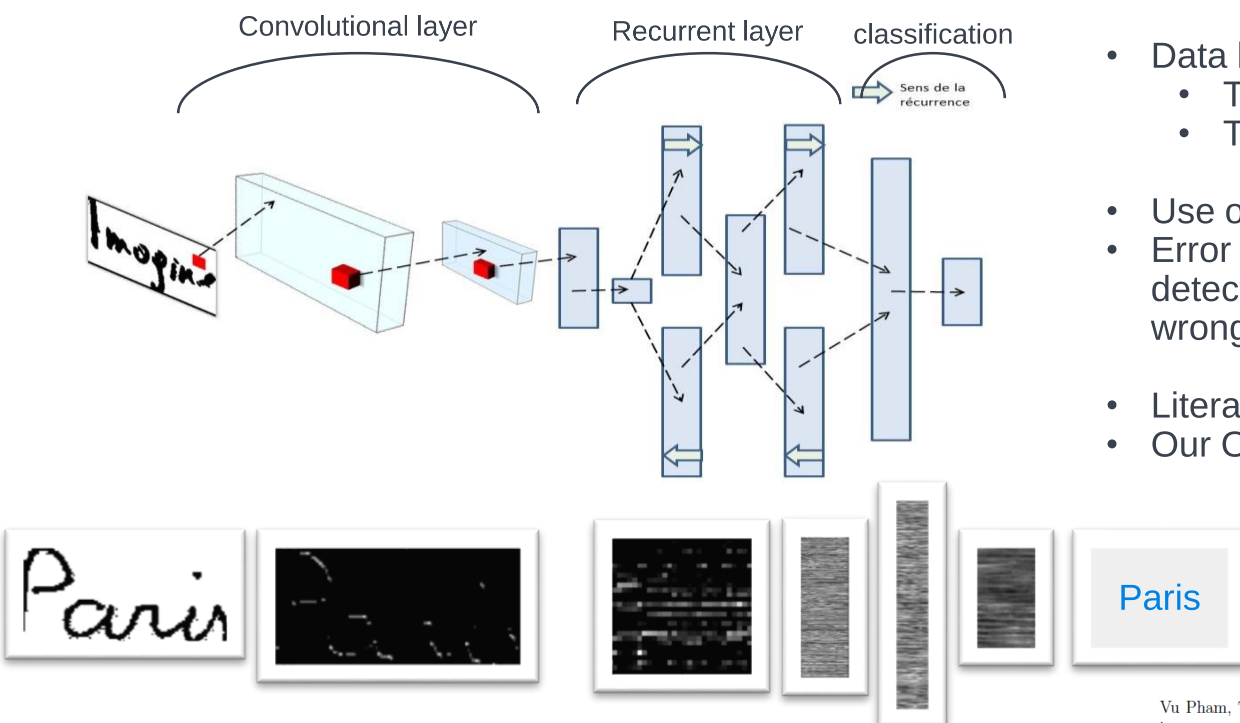
Different accuracies (results are based on claim reports and death certificates):

Scanned document: **99%+**
Scanned + handwritten: **61%** (literature 58-60%)

Scanned + handwritten can go up to **80%+** if we consider the context and prior on document.

OCR algorithms with handwritten writing style

Training and results



- Data base (Rimes)
  - Training on **70 000 pictures**
  - Testing 7000

- Use of AWS GPU
- Error metric WER: % of words badly detected (i.e. at least one character is wrong)

- Literature best score (LSTM 100): **44.37 %**
- Our OCR model: **46.5 %**

Vu Pham, Théodore Bluche, Christopher Kermorvant, and Jérôme Louradour. Dropout improves recurrent neural networks for handwriting recognition. In *Frontiers in Handwriting Recognition (ICFHR), 2014 14th International Conference on*, pages 285–290. IEEE, 2014.

Focus 3
Telematics

General overview

Telematics

At a glance (1/3)

- **Key words** within the insurance business:
  - Black Box Insurance
  - Usage Based Insurance
  - Pay As You Drive
  - Pay How You Drive
  - Driving pattern
  - Driving styles
  - Pattern recognition
  - Data provider
- The information is called driving behavior data (DBD). It covers the typical questions: when, where, how.
- **Disruption**
  - **Customers**: demand personalized relationships, reactivity, transparency, valuable services
  - **Insurers**: looking to boost profitability, anticipate and understand clients needs
  - Other **players**

Telematics

At a glance (2/3)

- Applications:
  - Pricing,
  - Scoring,
  - Claims (first notice of loss),
  - Fraud,
  - etc.
  - **Underwriting & Services**.
- Involve potential causative variables in the equation that could replace **traditional proxies**:
  - **Accuracy**
  - **Incentives**
- **Telematics is no longer in early stages:**
  - Italy, UK, US
  - French market

Telematics

At a glance (3/3)

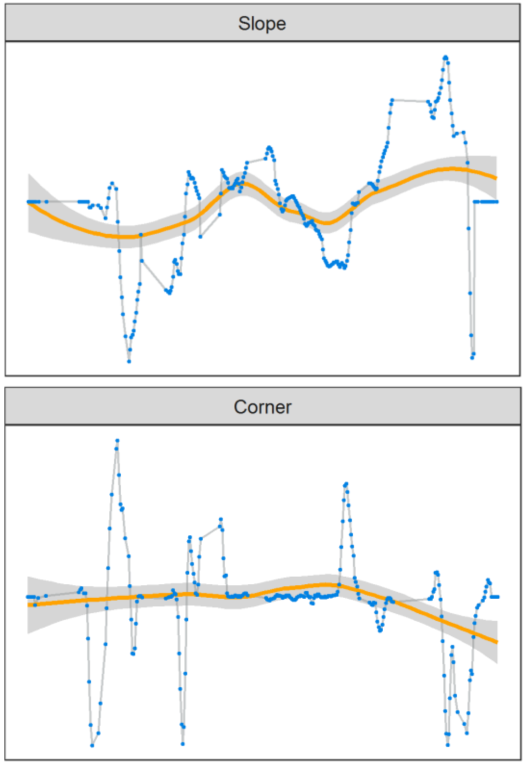
In practice

Telematics – in practice

Data preparation

- Provided by (at least) **GPS** latitude, longitude and **timestamp**.
- **Data management**: filtering techniques (causal moving mean or median, Kalman, etc.)
- Compute **indicators**: speed, acceleration, etc.

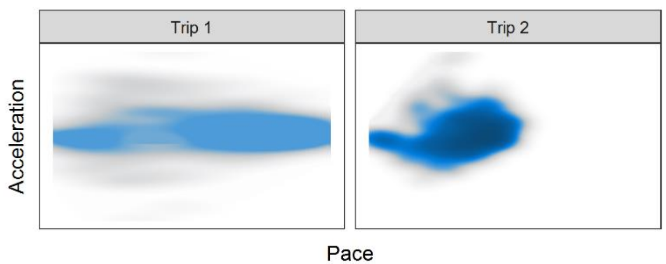
Latitude	Longitude	Elevation	Time
degree	degree	meters	dd/mm/yyyy hh:mm:ss
48.8755	2.2848	50	21/05/2018 20:30:00
48.8760	2.2868	55	21/05/2018 20:30:05
48.8757	2.2884	60	21/05/2018 20:30:10
48.8755	2.2905	65	21/05/2018 20:30:15
48.8753	2.2919	75	21/05/2018 20:30:20

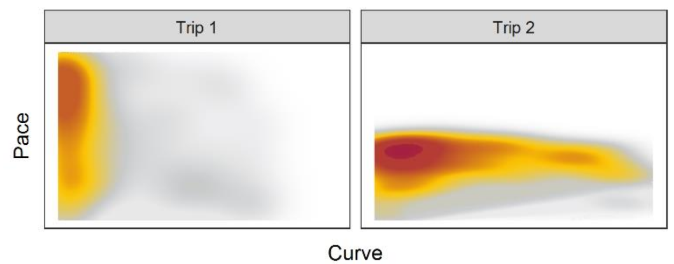


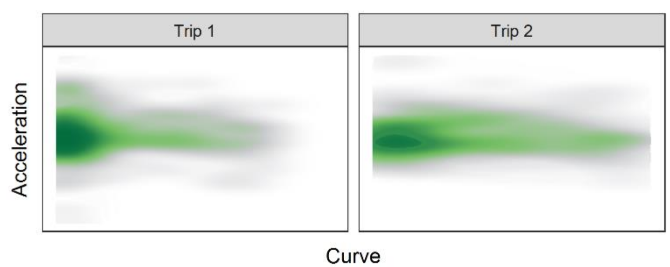
Telematics – in practice

Data analysis – overview

- Build your scores to answer questions:
  - How **far** do you drive?
  - How **aggressive** are you?
  - How **fast** do you drive?
  - **When** do you drive?
  - …







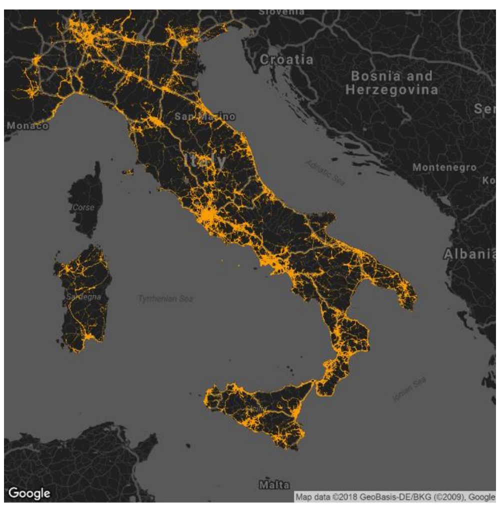
Telematics – Case study

Get insights from the data

- From very raw data, we can enrich/challenge the KPIs reported by the data provider. Information is being controlled with a lot of applications (**pricing** and **services**).
  - More: White paper 2018 http://www.milliman.com/uploadedFiles/insight/2018/raw-telematics-data-driving-behaviour.pdf

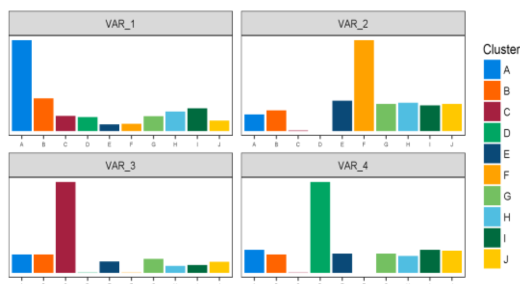
Data assessment
Take control

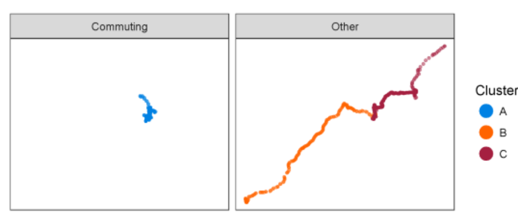
- **Explore** with simplified analysis
- **Scrub** the data
- According to the frequency we can **compute information**



Visualization
View the data

- **Use custom graphics** to take decisions for the modelling part





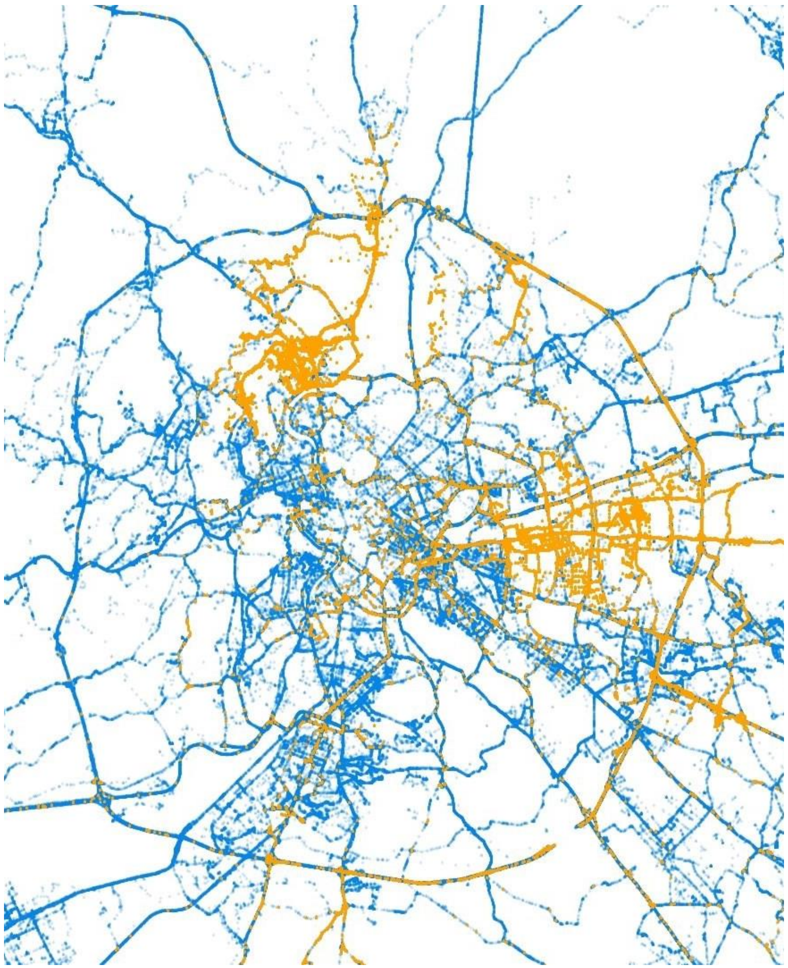
Use information
Build profiles, KPIs, services

Use the information for:

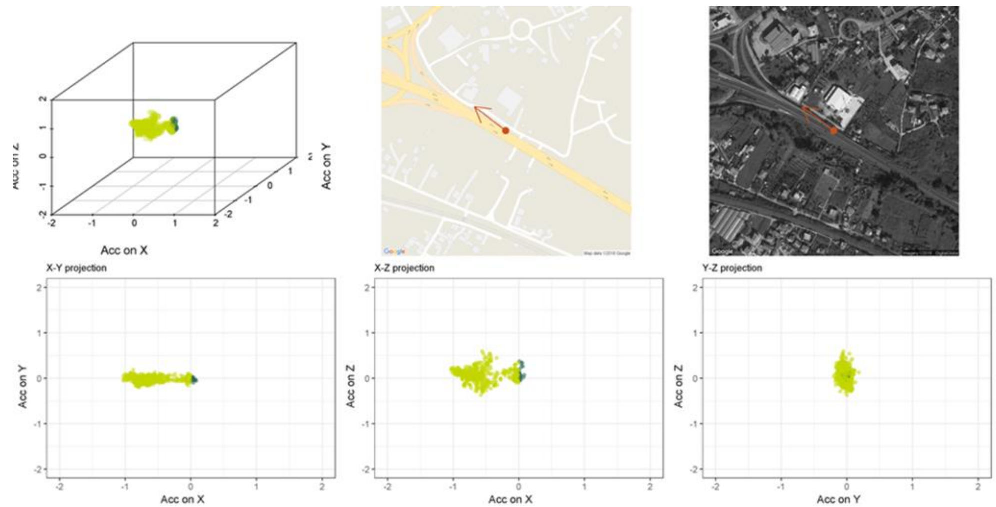
- **Trips and drivers analysis:** unsupervised techniques to build objective clusters
- **Crash analysis**: get a better understanding of how an accident happened
- **Build a driving risk score** the driver can use to optimize the insurance premium
- **Fraud detection:** check if claims report matches with the telematics data
- **Predictive maintenance**: considering the data available, predict and anticipate the maintenance of the vehicle
- **Etc.**: other roads conditions, etc.

Telematics – Case study

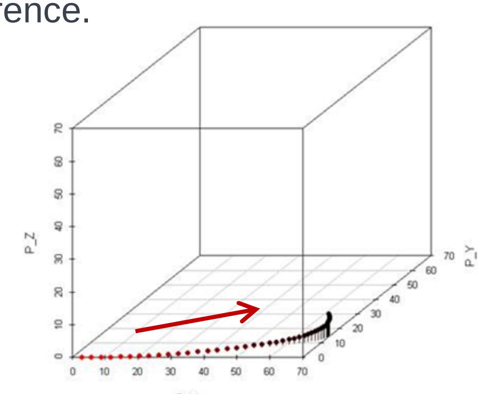
Focus: How to analyze crash/event?



- Analysis of detected events.



- Trajectory inference.



Conclusion

Thank you. Questions?

Milliman