

Natural Language Processing use cases in the insurance industry

Floriane MOY, Milliman

September 2020

About the speaker

Floriane MOY

- **Milliman** is a leading **actuarial** firm worldwide, with a successful track record of consulting to the insurance market.
- For more than 10 years, the diversity of consultants' profiles (actuaries, data scientists, data engineers) and their experience allow the **Advanced Analytics** team to work on a wide range of projects and ensure their success. At the forefront of innovation, the department is familiar with the latest technologies and highly proficient in modeling.
- **Floriane** is a Data Scientist Consultant in France and is a member of the Analytics department. She has extensive experience in statistics and machine learning, as well as in the use of open source technologies.

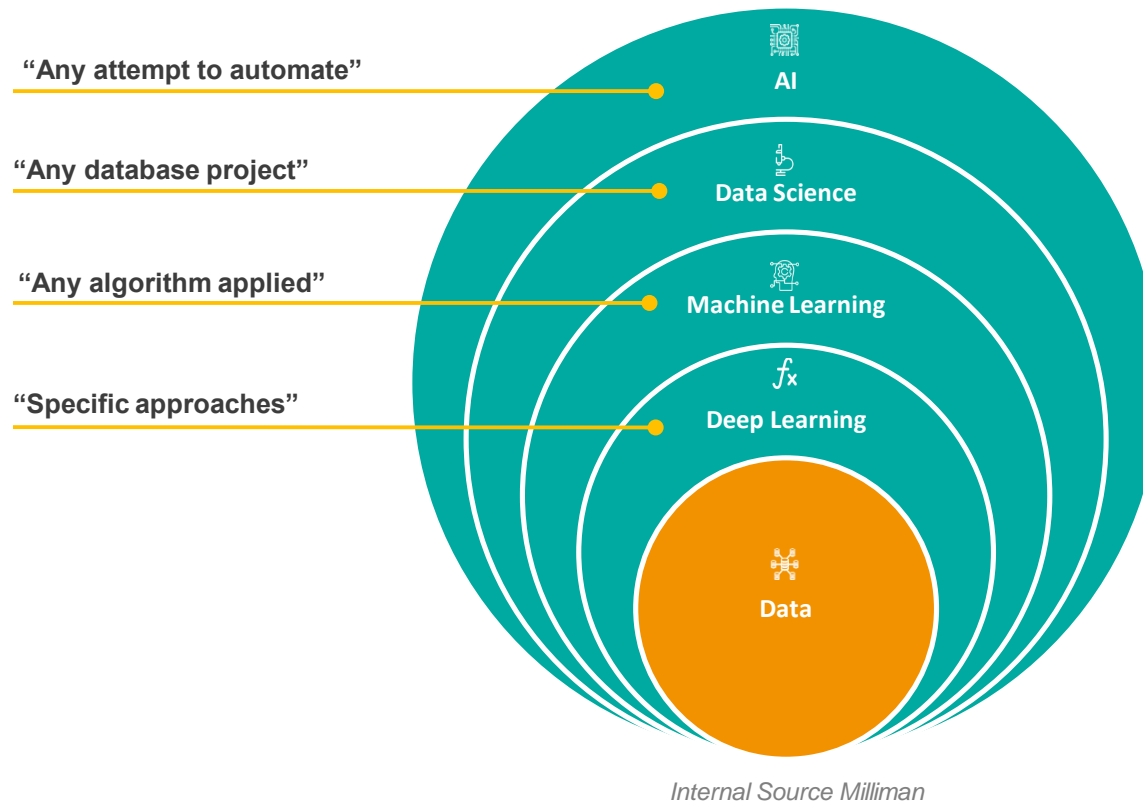


Agenda

- What is Natural Language Processing (NLP)?
- NLP within the insurance industry
- Use cases
 - Topic classification
 - Sentiment analysis
 - Text generation

What is NLP?

A science of language



- We present here a schematic representation of **AI concepts**.
- From 1950 to today, NLP has evolved:
 - Conceptual ontologies
 - Humanlike robots
 - Basic computing tasks
 - Machine Learning
 - Deep Learning
 - Etc.
- NLP is an extremely vast field, at the **frontier** of different notions, from the most trivial analysis to the most complex concepts (deep learning).

What is NLP?

Different approaches and methods (1)

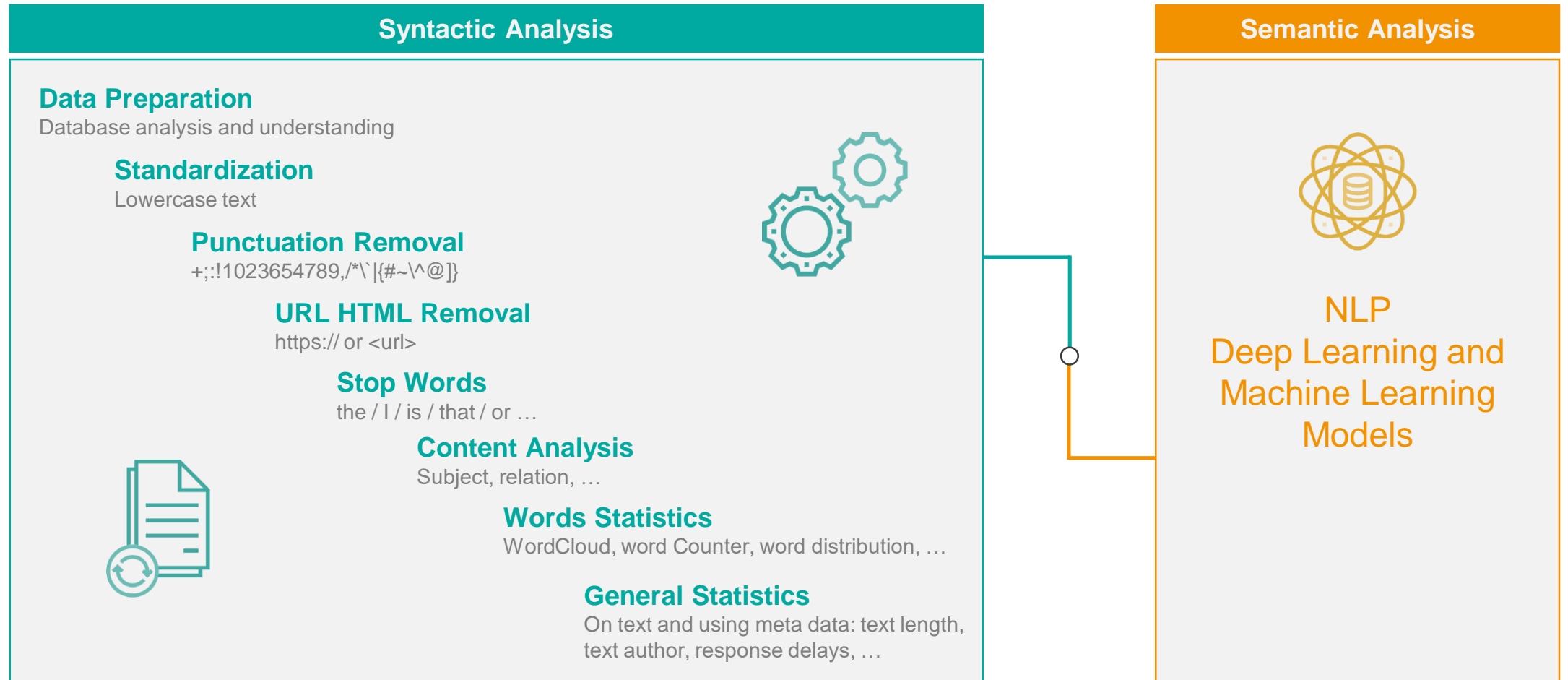
- NLP methods can be separated into two main groups:

Syntactic Analysis	Semantic Analysis
Tokenization	Name Entity Recognition
Stop Words	Word Sense Disambiguation
Stemming	Natural Language Understanding
Bag of Words ...	Natural Language Generation ...

- **Syntactic Analysis** is mainly used for basic text processing, while **Semantic Analysis** - often involving machine learning models - is used to go further and try to infer deeper meaning.

What is NLP?

Different approaches and methods (2)

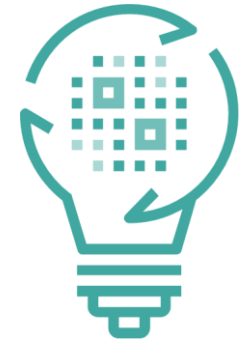


Internal Source Milliman

NLP within the insurance industry

Opportunities

- All players in the insurance market today are looking to **improve their internal processes**. This includes streamlining and accelerating procedures that can be automated to increase efficiency and focus more attention on matters of importance.
- **All topics are affected**, from underwriting to claims payment, not to mention pricing, reserving or regulatory requirements.
- If the use of the supervised learning machine is already well anchored among insurers to improve their models, the democratization of open source and the significant growth of APIs is gradually bringing elements of artificial intelligence into the daily lives of actuaries.
- **NLP is particularly becoming more and more widespread** and offers new perspectives to insurers seeking innovation and cost optimization.



NLP within the insurance industry

Data

- **Text-based information** is already widely available within the insurance industry.

TYPE

- PDFs
- Words and PowerPoints
- Internal and external e-mails
- Messages from chats
- Scanned documents
- ...



CONTENT

- Underwriting information
- Contracts
- Technical specifications
- Claims details
- Clients requests
- Online reviews and feedbacks
- Internal and external reports
- ...

- However, this information is not always fully explored, nor streamlined in more general processes.
- NPL offers new possibilities to **efficiently process** this information.

Use cases

Examples of what NLP can bring to the insurance industry

- We present hereafter a non-exhaustive list of some **promising examples**.



Email sorting



Claim orientation



E-reputation



Client feedback



Document Filling



Automated response

Use cases

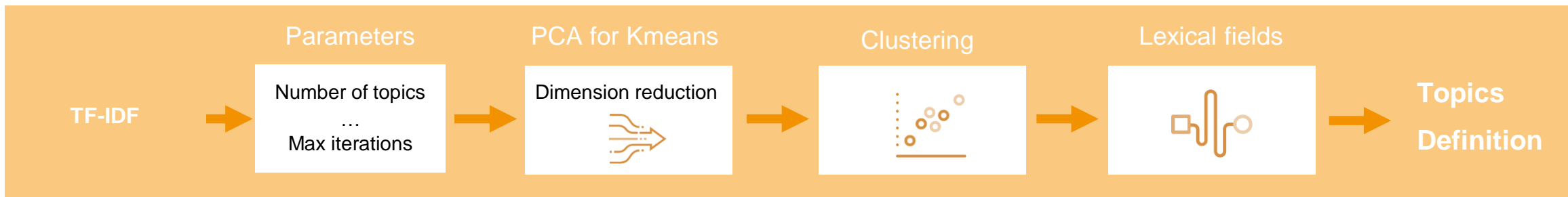
Topic classification (1)

- Classification into subjects allows the identification of general themes within a corpus of texts.
 - Example of how computers can understand human language: the **TF-IDF** matrix.



Internal Source Milliman

- K-Means** and **LDA**: unsupervised methods for Topic classification.



Internal Source Milliman

Use cases

Topic classification (2)

- Implementing topic classification algorithms is relevant to many **different situations**, in order to help to handle **large volumes** of data or to provide a **reading grid**.
- Below is an overview of results that illustrate this point:

Sorting clients requests

With several emails received each day, it is interesting to be able to sort the information in order to process it more efficiently.

TOPIC 1 MOTOR	TOPIC 2 HOUSEHOLD	TOPIC 3 GENERAL
coverage	house	information
pay	apartment	contract
vehicle	problem	quote
change	tenant	date
driver	address	wonder
share	fire	increase
car	family	problem
park	move	call

Customer relationships

Many comments are available on the internet, it may be useful to carry out a study in order to draw some conclusions.

Topic 1: Too expensive

It's worth nothing, I've been waiting for the refund for months. I advise you to run away. It is way too expensive (beware of accidents abroad). I don't even mention the legal protection which is useless.

Topic 2: One bad experience

In June 2017, I had a water infiltration at my place. They first told me I was covered and would reimburse me 250€ per tarpaulin, but they didn't reimbursed me anything. And I even got a malus when I was guaranteed not to.

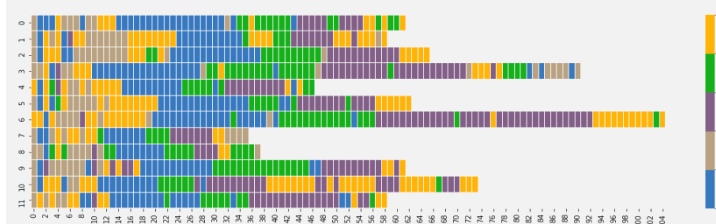
Topic 3: Service provided

Although slightly more expensive than many of its competitors, this insurer has demonstrated excellent coverage in the event of an automobile claim.

Reports analysis

Some reports are sometimes difficult to apprehend. A reading grid can help to digest the information and guide the reader to find relevant information more rapidly.

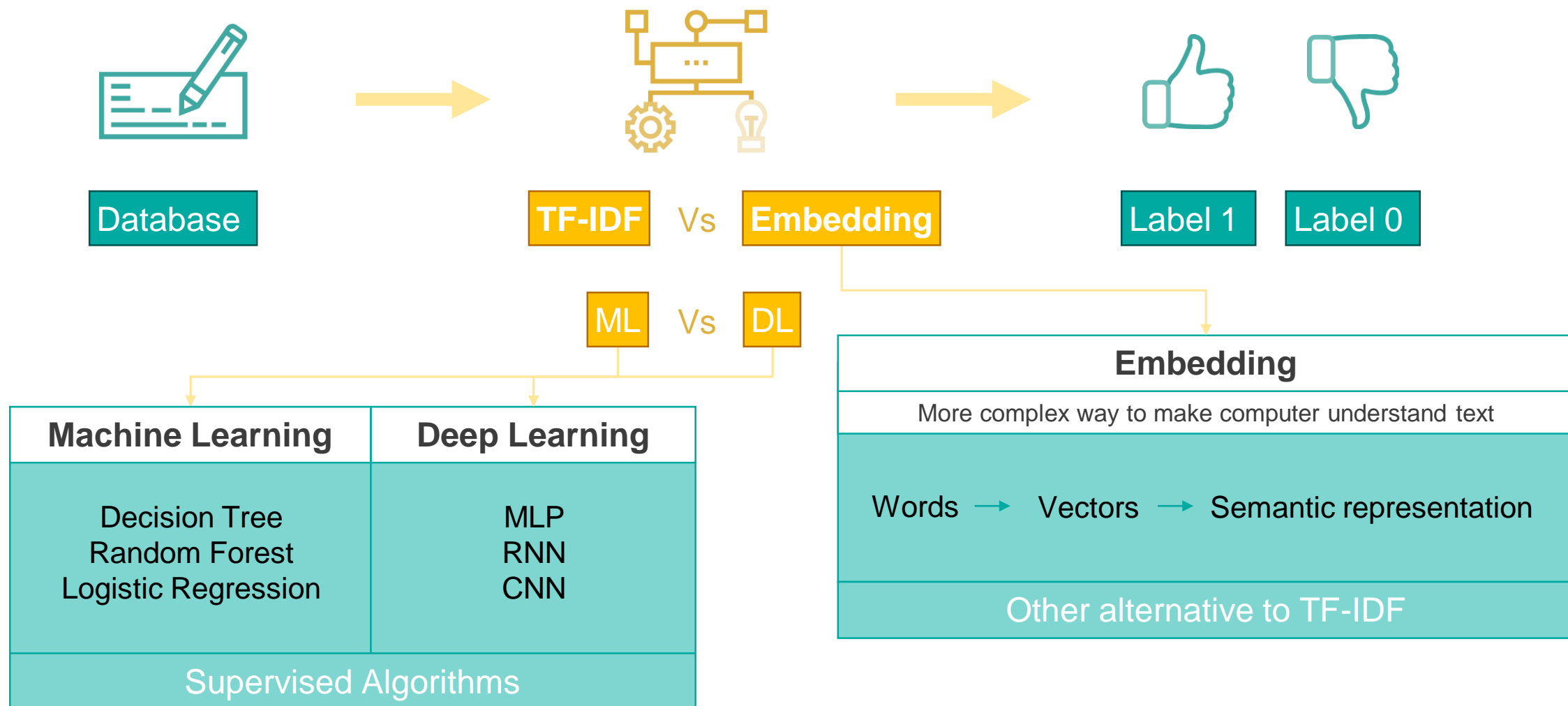
Transition of dominant topics on 12 PDF files and pages (training data) with number of topics = 5.



1. Management, risk, board, committee, ...
2. Report, financial, solvency, information, audit, ...
3. Liability, asset, valuation, solvency, ...
4. Risk, market, model, exposure, credit, ...
5. Capital, business, fund, scr, ...

Use cases

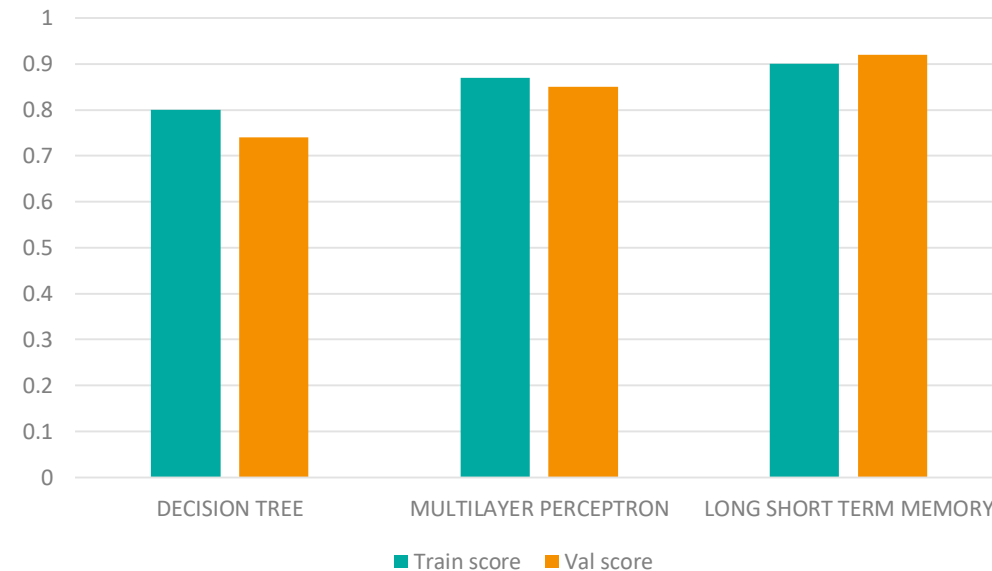
Sentiment analysis (1)



Use cases

Sentiment analysis (2)

- A first **database with labels** available is necessary in order to **train a model**: this is typically done by splitting such database into a training sample, a validation sample and a test sample.
- The label is helping the algorithm to learn which text is associated with a **positive** (label 1) and a **negative** sentiment (label 0).
- Three models trained:
 - A **decision tree** (CART model)
 - A **multilayer perceptron**
 - A **long short term memory** neural network
- These models can be easily compared to already trained models that exist in open source libraries:
 - Sentiment Analyzer from the **NLTK** Python library



Use cases

Sentiment analysis (3)

■ Test on specific texts:

Prediction > 0.5 means positive

Prediction < 0.5 means negative

My app has been giving me an error code and failing to update for the past week.

▼ LSTM: 0.082

▼ MLP: 0.065

▼ NLTK: 0.126

I have been locked out of my account because of incorrect password entry. I can provide details if needed.

▼ LSTM: 0.124

▼ MLP: 0.068

► NLTK: 0.5

I can't use my account, no phone or contact options are available, so I used the live chat.

▼ LSTM: 0.245

▼ MLP: 0.197

▲ NLTK: 0.778

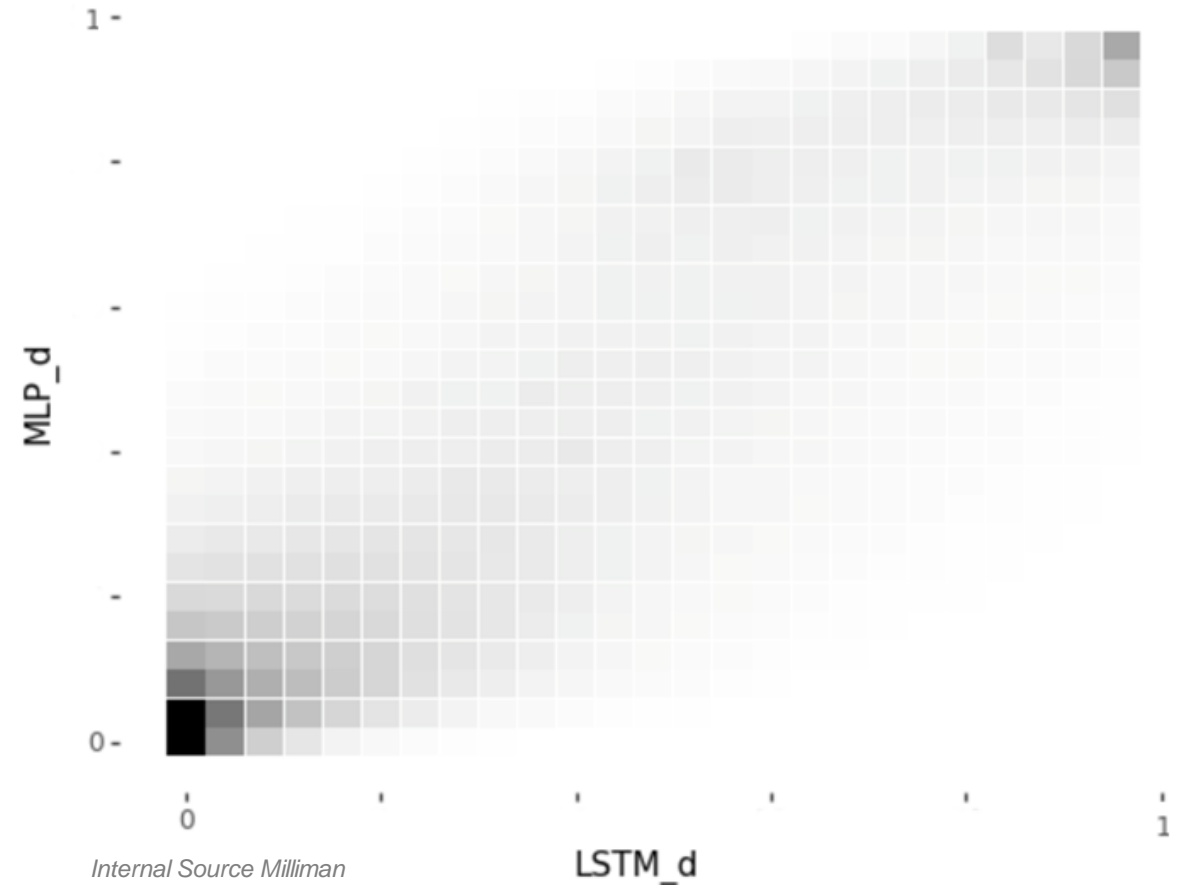
Very good support thank you for your help!

▲ LSTM: 0.988

▲ MLP: 0.920

▲ NLTK: 0.966

LSTM vs MLP vs NLTK Prediction

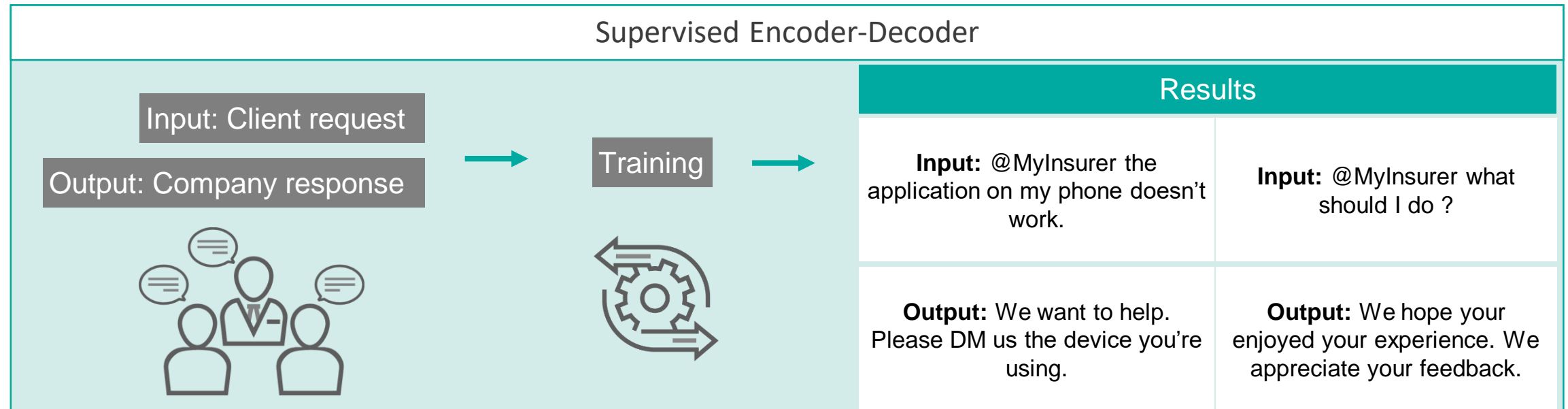


LSTM vs MLP prediction

Use cases

Text Generation

- Below is a non-exhaustive list of algorithms which can understand how text is written:
 - N-gram: Can learn words frequencies to anticipate probabilities of word occurrences.
 - Recurrent Neural Networks (RNN) LSTM: Can learn with past information (text previously available).
 - Encoder / Decoder: Can proceed when input and output lengths are different (e.g. online translator).
 - GPT-2 API: New generation of multi-task algorithms.



Conclusion

- Our results:
 - Our concrete use cases show promising results for the insurance industry and are replicable on large databases.
 - Unsupervised **topic classification** algorithms work very well and present different applications.
 - **Sentiment analysis** models are robust and can be refined more precisely for a specific use case by training the models using a labeled database.
 - The main challenge for **text generation** is probably the computation time needed for algorithms trainings.
- Further explorations for the insurance industry:
 - How to automate the filling of documents?
 - How to evaluate message with high and low priority responses?
 - How to deal with unsupervised database for sentiment analysis?
 - How to exploit other text database available?
- Constant innovations are shared across the AI community. For instance, **GPT-3** is a recent new text processing model which comes from OpenAI.

Milliman – Advanced Analytics Department

Floriane MOY – floriane.moy@milliman.com

Thanks to the members of the Milliman Analytics team, particularly Rémi Bellina and Timothée Robert, for their contributions.

