Cyber incident reports: extrapolating severity using neural networks

Hugo Rapior, Detralytics

5th European
Congress of Actuaries
www.eca2024.org

About the speaker

- **Rapior Hugo –** *TCP Consultant, Detralytics*
  Hugo is part of the TCP program at Detralytics, and works on R&D topics, P&C pricing, modelling and more especially on cyber.

Detralytics is a consulting and training firm focused on Actuarial Science, Data Science and Risk Management. Detralytics was founded to support companies in the advancement of actuarial science and help them to solve their challenges, while offering the perfect stepping stone to the most talented actuarial students. We guarantee to go beyond traditional consulting, by offering a unique combination of academic expertise, deep career and market knowledge.

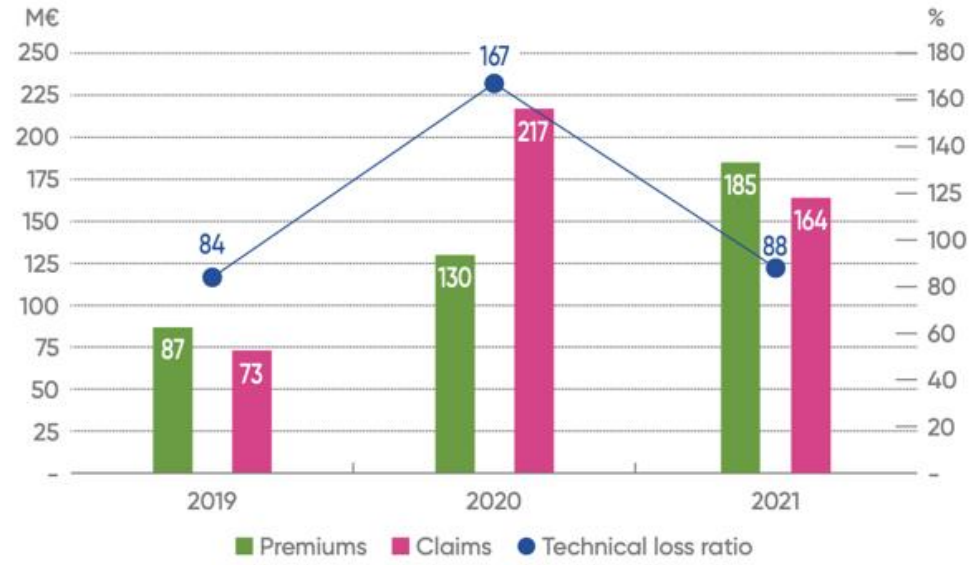
Introduction

- LUCY Report from the AMRAE :

- Loss ratios : 84% in 2019, 167% in 2020, 88% in 2021
- Premiums : +44% collected, for a growth in numbers o 27,5%

- Report from the General Treasury Directorate on the cyber risk, in September 2022:

- Raises questions on data
- Points the necessity of innovating methods (e.g. the bayesian)



Source: 2022 AMRAE LUCY Study.

Introduction

- **Question** : Can we use the multiple textual ressources at hand (such as the incident/claims descriptions) to get a better understanding of the underlying risk?

- **Objectives** :
  - Data augmentation
  - Transform "literary" insight in quantitative insight, to feed into bayesian models
  - Better anticipate and foresee within the processes of incident management

Summary

1- Embedding techniques

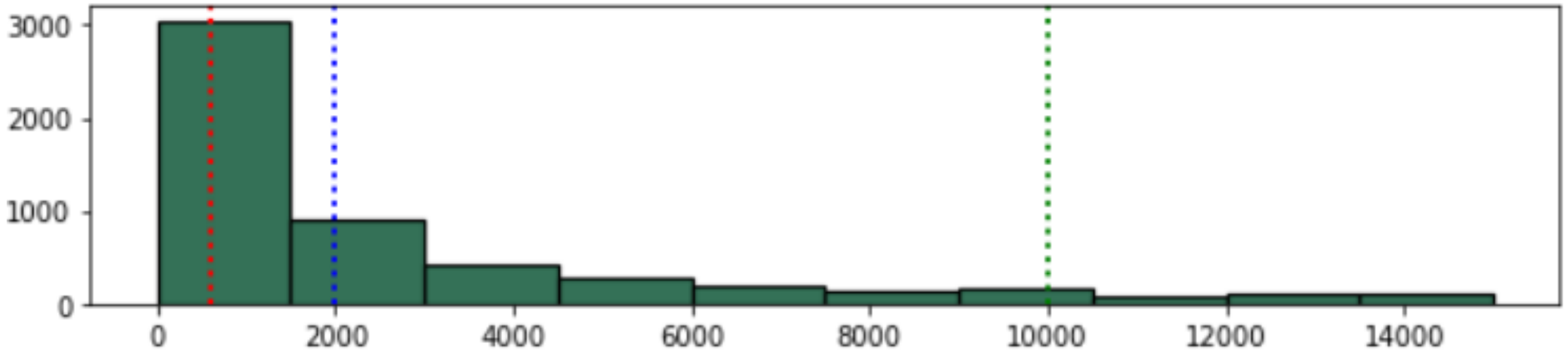
2- Neural Networks

3- Model enrichment

4- Uses and perspectives

Base PRC: Privacy Rights Clearinghouse (US)

- Variable being a marker of severity (the number of records)and the claim description



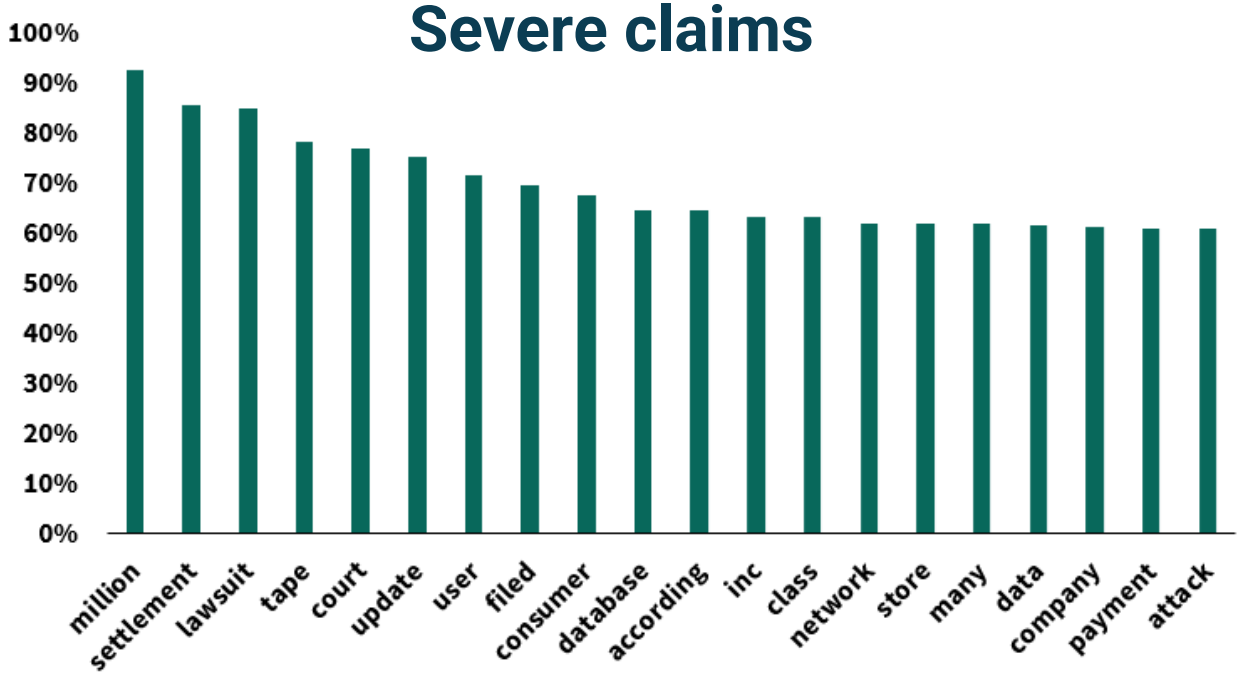
- Splitting of the base and flagging of the most severe claims using the number of records variable.

NB OF RECORDS < 4767 ?	NB OF RECORDS > 4767 ?
Attritionnal claim	Severe claim

Base PRC: Privacy Rights Clearinghouse (US)

- The words used in the claim description also inform on the level of severity

Severe claims



Attritionnal claims

The following words are associated with the decription of attritionnal claims :

- Paper
- Document
- Dishonest
- Accidentally
- School

Pre-treating the textual data

Treating the claim description

Claim description	Cleaned description
Union Hospital suffered an inadvertent disclosure on approximately 1/18/16 that resulted in 1 record being exposed, which included social security numbers.	union hospital suffered inadvertent disclosure approximately resulted record exposed included social security number

Interfering information :

- Stopwords
- Dates
- Punctuation
- Numbers

Pre-treating the textual data

Treating the claim description

Claim description	Cleaned description
Union Hospital suffered an inadvertent disclosure on approximately 1/18/16 that resulted in 1 record being exposed, which included social security numbers.	union hospital suffered inadvertent disclosure approximately resulted record exposed included social security number

Interfering information :

- Stopwords
- Dates
- Punctuation
- Numbers

Corpus dictionary

- Each word within the corpus is called a **token**
- An analysis of the text allows the identification of new tokens, each one having 2 or 3 words to add in the dictionnary

Pre-treating the textual data

Treating the claim description

Claim description	Cleaned description
Union Hospital suffered an inadvertent disclosure on approximately 1/18/16 that resulted in 1 record being exposed, which included social security numbers.	union hospital suffered inadvertent disclosure approximately resulted record exposed included social security number

Interfering information :

- Stopwords
- Dates
- Punctuation
- Numbers

Corpus dictionary

- Each word within the corpus is called a **token**
- An analysis of the text allows the identification of new tokens, each one having 2 or 3 words to add in the dictionnary

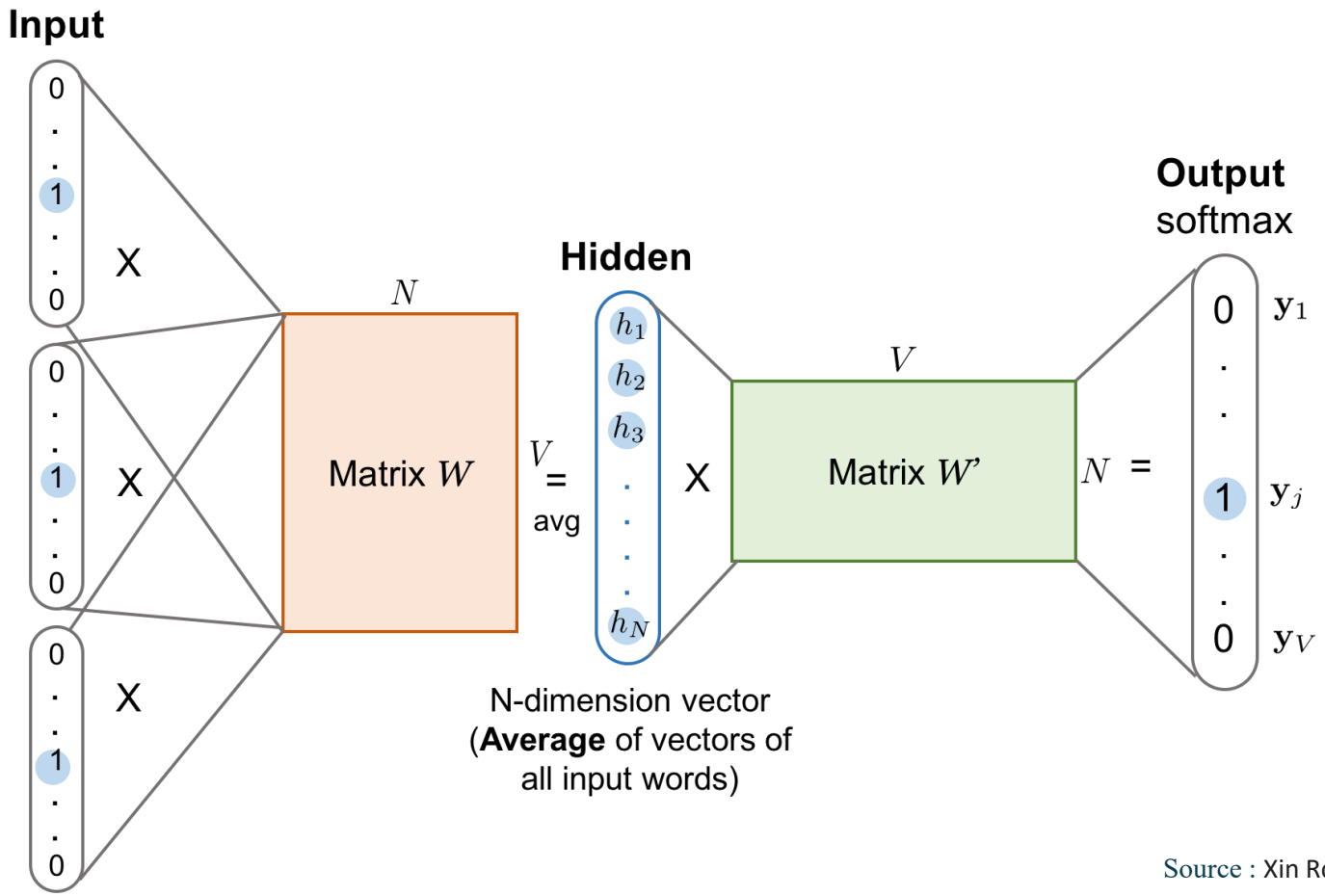
Relevant string of words

« social security number »

« personal information »

« email adress »

Word embedding – Word2Vec



Source : Xin Rong. word2vec Parameter Learning Explained

Word embedding – Word2Vec

A word is encoded in an **N-dimensional space**

Words with close **meaning** or **influence** will be close in that space

	Man (5182)	Woman (9742)	King (4815)	Queen (7464)	Apple (421)	Orange (6151)
Gender	-1	1	-0,95	0,97	0,00	0,01
Royal	0,01	0,02	0,93	0,95	-0,01	0,00
Age	0,03	0,02	0,7	0,69	0,03	-0,02
Food	0,04	0,01	0,02	0,01	0,95	0,97
…	…	…	…	…	…	…

Word embedding – Word2Vec

A word is encoded in an **N-dimensional space**

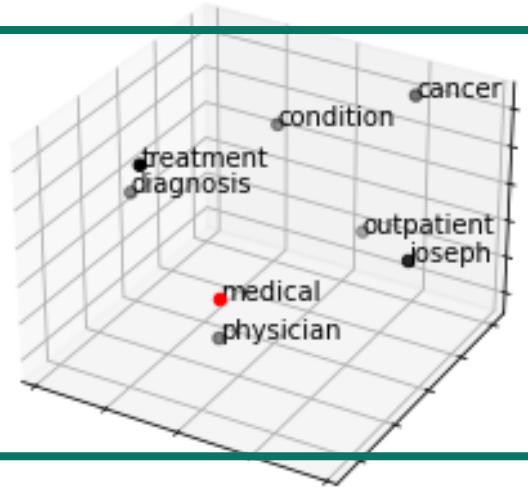
Words with close **meaning** or **influence** will be close in that space

Close words to « insurance »
Life
Insurer
Coverage
Enrollee
Plan
Guarantor
Aflac

	Man (5182)	Woman (9742)	King (4815)	Queen (7464)	Apple (421)	Orange (6151)
Gender	-1	1	-0,95	0,97	0,00	0,01
Royal	0,01	0,02	0,93	0,95	-0,01	0,00
Age	0,03	0,02	0,7	0,69	0,03	-0,02
Food	0,04	0,01	0,02	0,01	0,95	0,97
…	…	…	…	…	…	…

3D Representation

The close words to « **medical** » in our corpus of claim description can be **3D plotted**



Summary

1- Embedding techniques

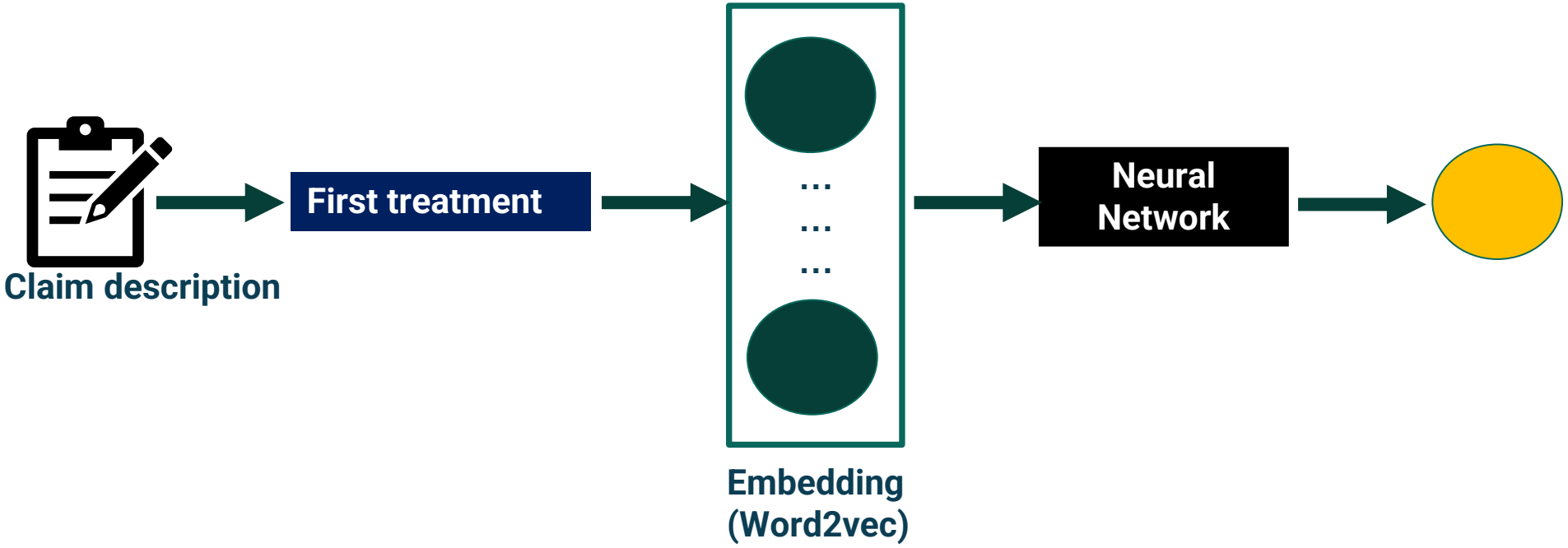
2- Neural Networks

3- Model enrichment

4- Uses and perspectives

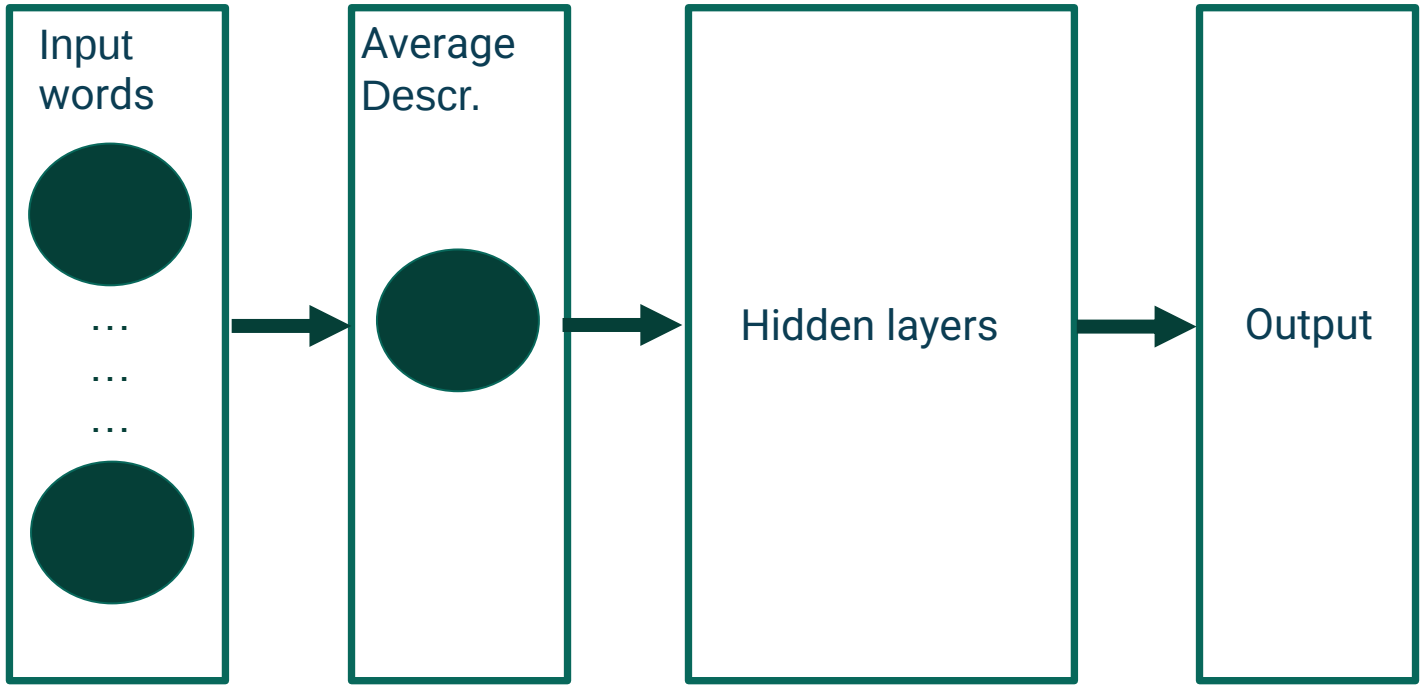
Summary of the method

Methodology



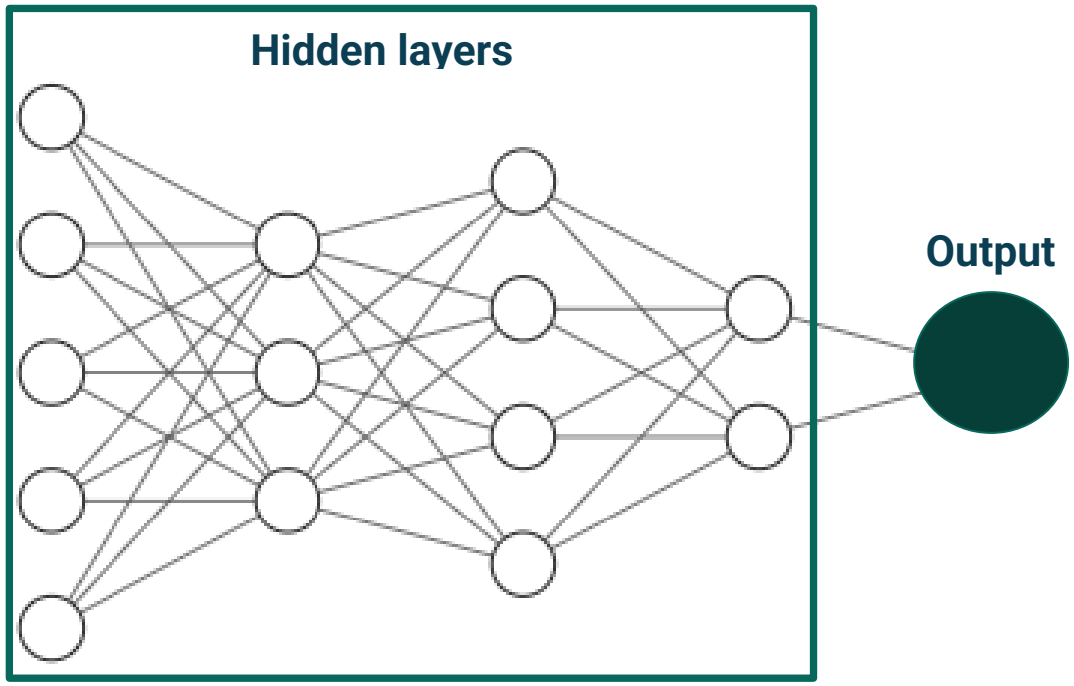
The family of model we chose: Multilayer perceptron

- Zooming on the part **Neural Network** -> Multilayer perceptron



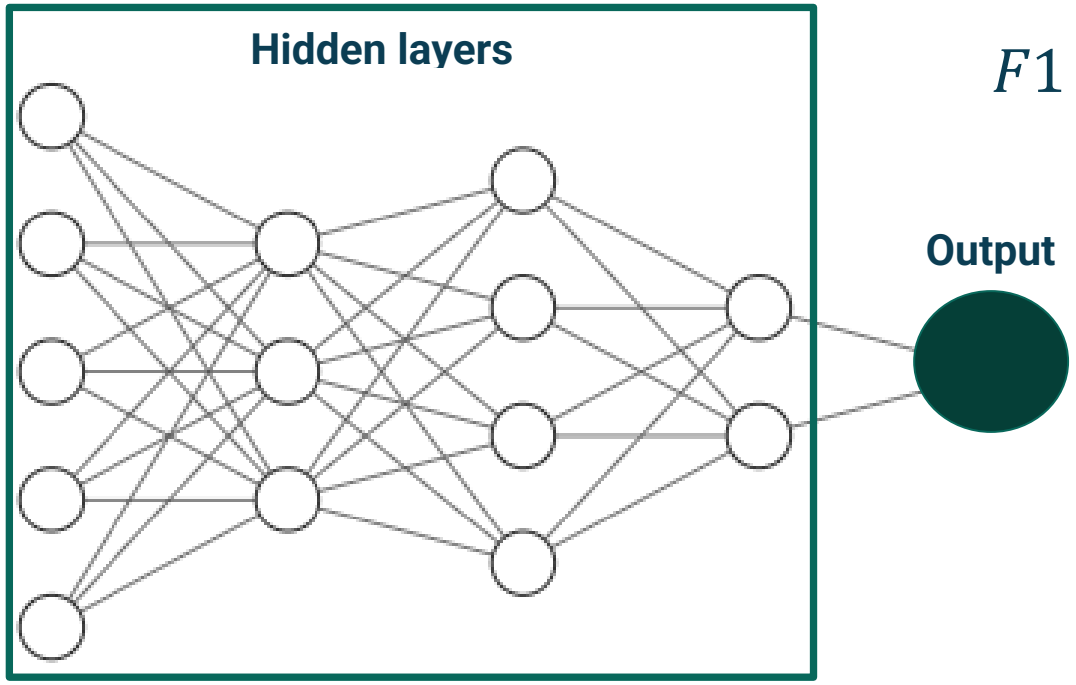
The family of model we chose: Multilayer perceptron

- Gridsearch on the hidden layers of the neural network (F1 score)



The family of model we chose: Multilayer perceptron

- Gridsearch on the hidden layers of the neural network (F1 score)



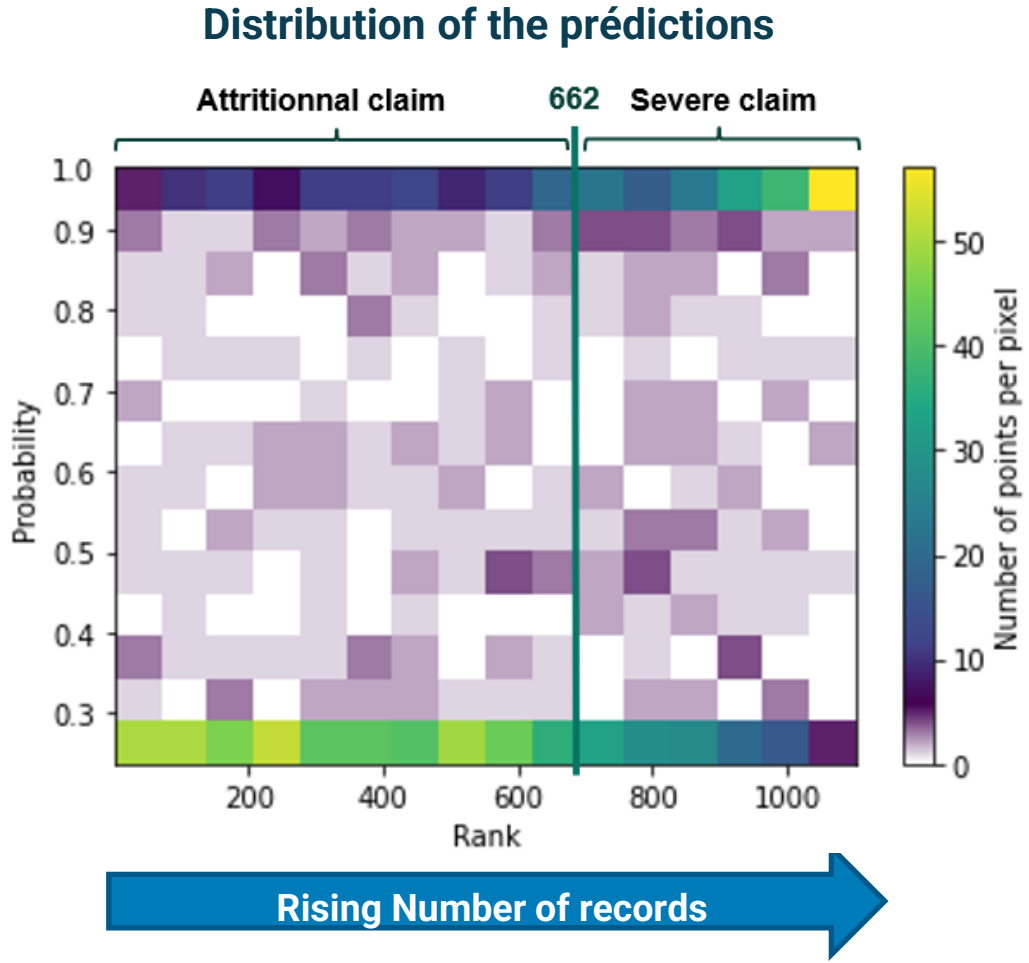
$$F1\ score = \frac{TP}{TP + MEAN(FP; FN)}$$

TP : True positives
FP : False positives
FN : False negatives

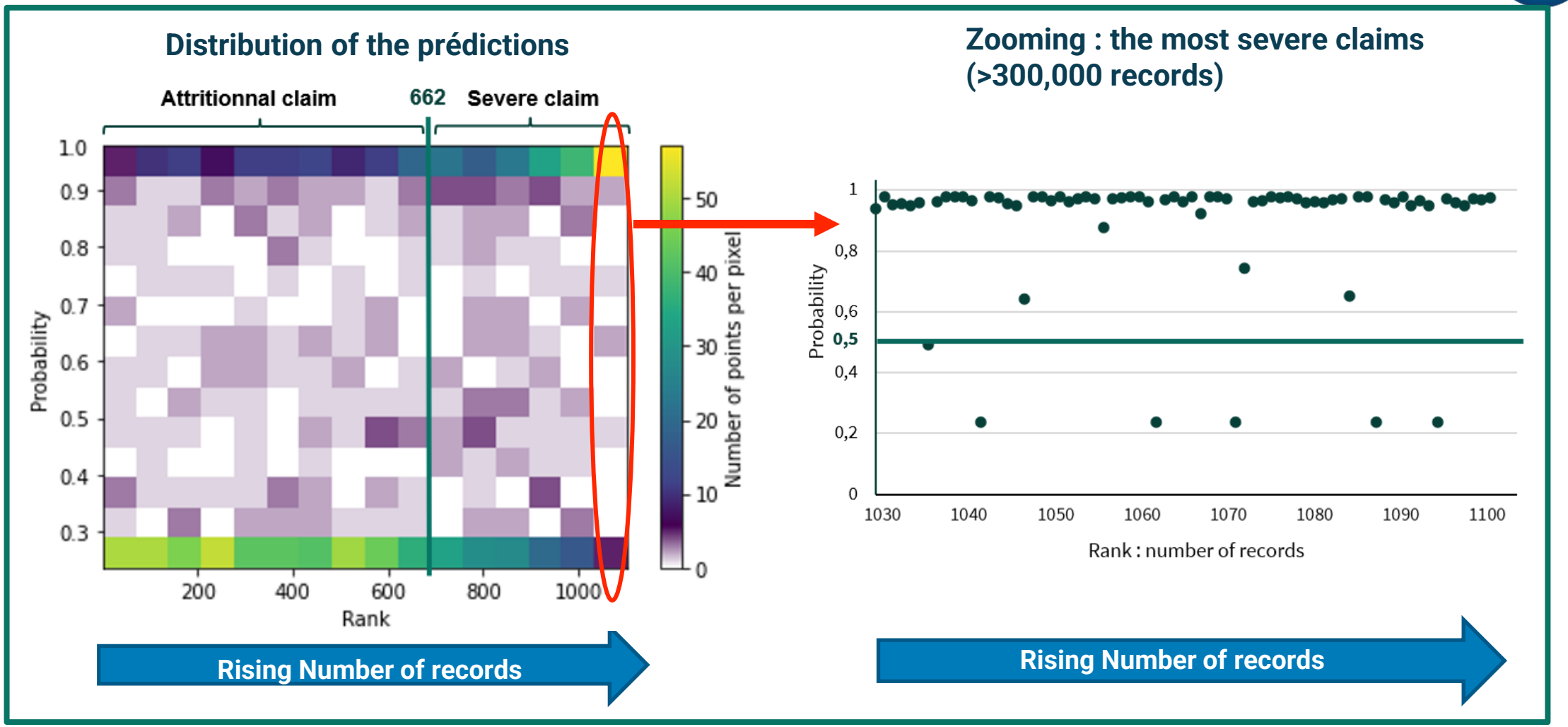
		Severity?	
		0	1
Pred	0	479	176
	1	182	265

F1-Score = 60%

Results : Distribution of the predictions



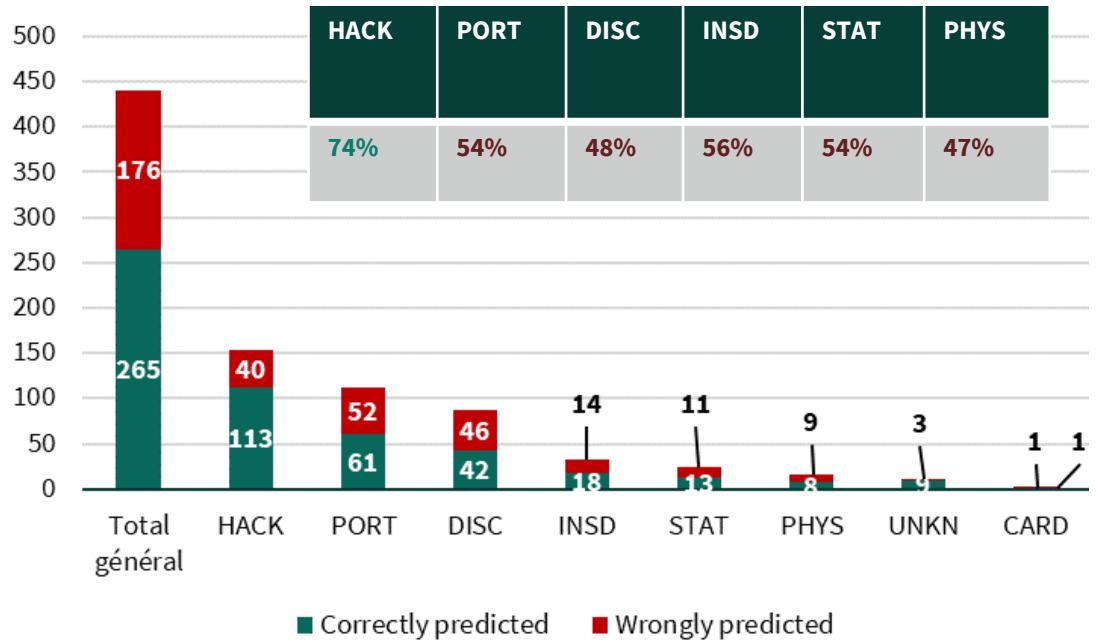
Results : Distribution of the predictions



Prediction quality by type of intrusion

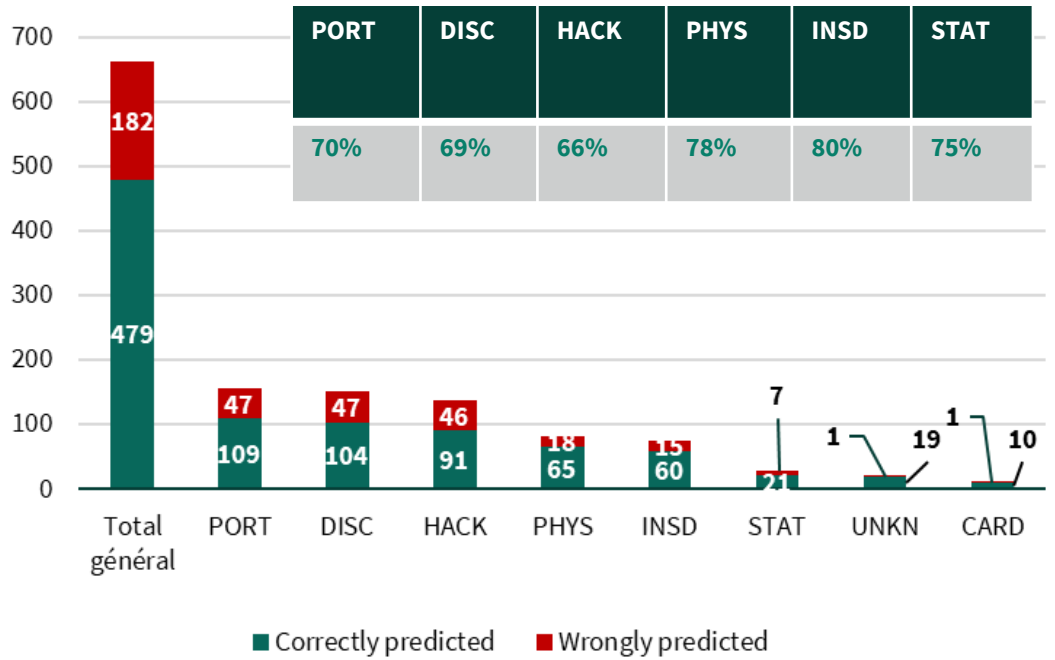
Severe claims prediction

HACK	PORT	DISC	INSD	STAT	PHYS
74%	54%	48%	56%	54%	47%



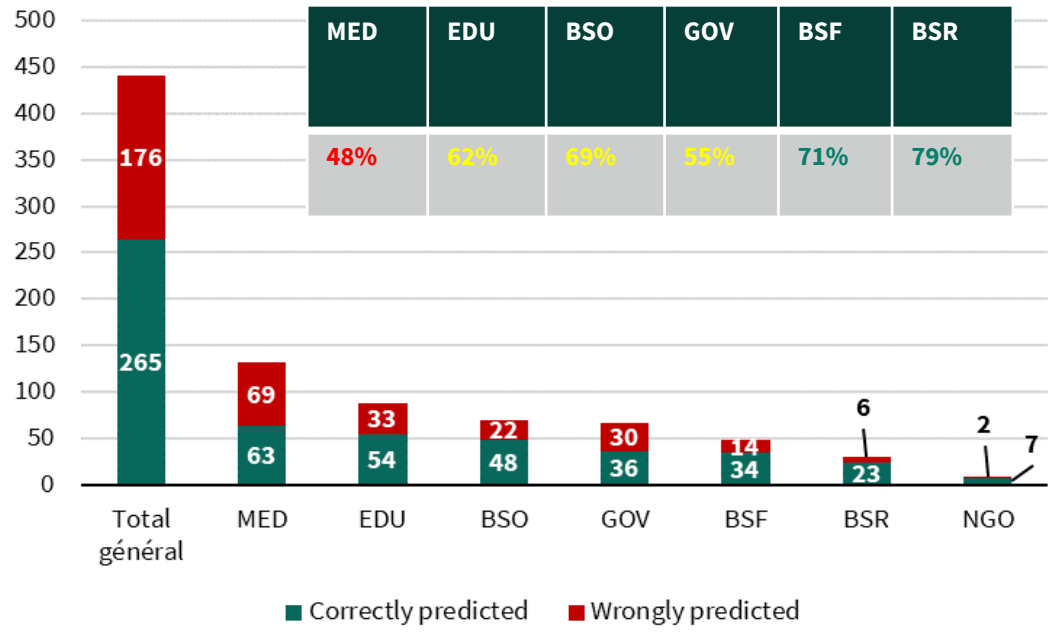
Attritionnal claims prediction

PORT	DISC	HACK	PHYS	INSD	STAT
70%	69%	66%	78%	80%	75%

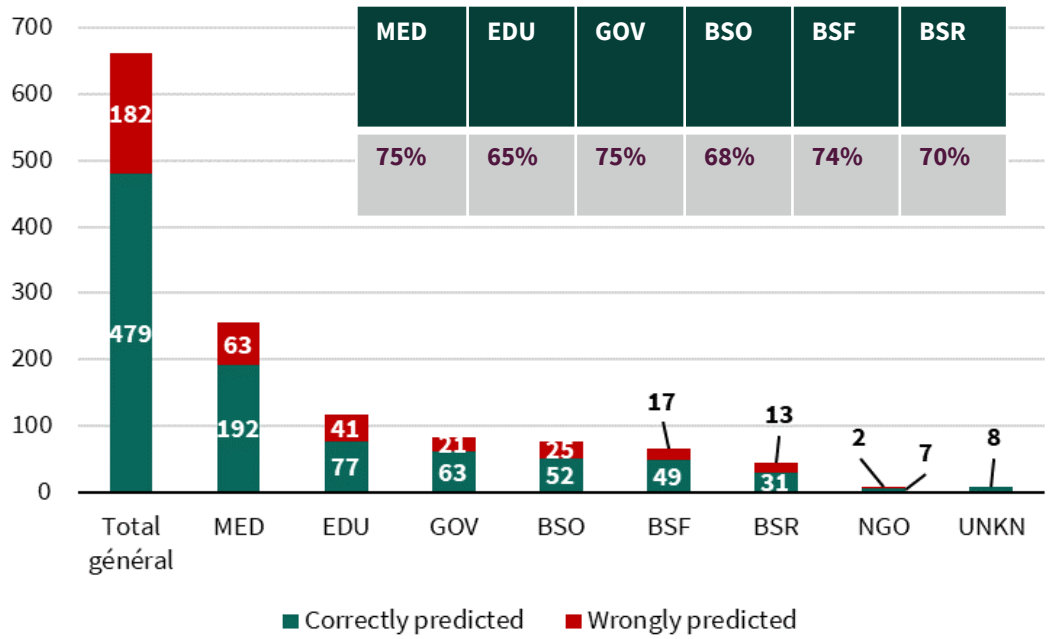


Prediction quality by type of organisation

Severe claims prediction



Attritionnal claims prediction



Summary

1- Embedding techniques

2- Neural Networks

3- Model enrichment

4- Uses and perspectives

Text mining : expressions régulières

A neural network cannot understand nor interpretate the link between numbers and words

12 social security numbers ≠ 12 companies ≠ 12 million

How to use and interpretate numerical data within claim descriptions ?

Text mining : expressions régulières

A neural network cannot understand nor interpretate the link between numbers and words

12 social security numbers ≠ 12 companies ≠ 12 million

How to use and interpretate numerical data within claim descriptions ?

- In 50% of the descriptions, we can observe frequent forms such as :

number × words

- **Those informations are directly linked to the value of the *number of records* variable**

=> The numerical data is a direct indicator of the claim severity

Text mining : expressions régulières

A neural network cannot understand nor interpretate the link between numbers and words

12 social security numbers ≠ 12 companies ≠ 12 million

Approximation :

*A hacker […] has potentially revealed the names, Social Security numbers, and, in some cases, the birth dates and bank accounts of **27,000** employees working at **1,900** companies nationwide.*

How to use and interpretate numerical data within claim descriptions ?

- In 50% of the descriptions, we can observe frequent forms such as :

number × words

- **Those informations are directly linked to the value of the *number of records* variable**

=> The numerical data is a direct indicator of the claim severity

Regular expressions

Dictionary

Word	nb. Pred false	nb. Pred True	Total	Predictive power ?
patient	16	55	71	77%
people	15	52	67	78%
million	43	5	48	10%
record	2	27	29	93%
student	4	24	28	86%
employee	2	22	24	92%
current	7	14	21	67%
individual	2	19	21	90%
year	12	7	19	37%
customer	3	15	18	83%
…	…	…	…	…

Word	nb. Pred false	nb. Pred True	Total	Predictive power ?
companies	5	0	5	0%

Approximation :

*A hacker […] has potentially revealed the names, Social Security numbers, and, in some cases, the birth dates and bank accounts of **27,000** employees working at **1,900** companies nationwide.*

Dictionnaire $\boldsymbol{V}$ et paires $(\boldsymbol{N_u}, \boldsymbol{M_u})$

$$\sum_{k=0}^{n} \boldsymbol{N_u 1_{M_u \in V}} < \boldsymbol{Seuil}$$

Let’s come back to our example

*A hacker […] has potentially revealed the names, Social Security numbers, and, in some cases, the birth dates and bank accounts of **27,000** employees working at **1,900** companies nationwide.*

27000*1 + 1900*0 = 27000 > 4700

This is a severe claim for our new estimator !

Let’s come back to our example

*A hacker […] has potentially revealed the names, Social Security numbers, and, in some cases, the birth dates and bank accounts of **27,000** employees working at **1,900** companies nationwide.*

27000*1 + 1900*0 = 27000 > 4700

This is a severe claim for our new estimator !

Type of organisation	Perceptron	Regular expressions
MED	63	+ 10
EDU	54	+ 11
BSO	48	+ 7
GOV	36	+ 10
BSF	34	+ 1
BSR	23	+ 0
NGO	7	+ 1

Number of severe claims with method used

Summary

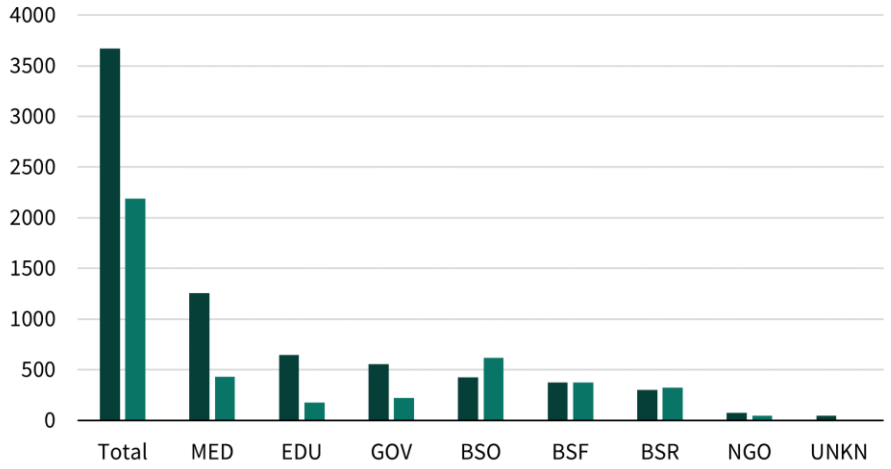
1- Embedding techniques

2- Neural Networks

3- Model enrichment

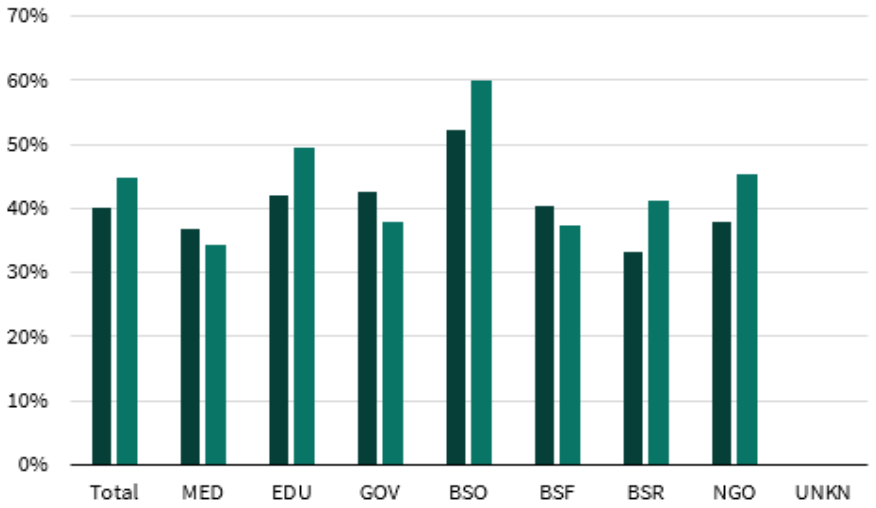
4- Uses and perspectives

Database : missing number of records



BSO, BSF, BSR

- The business categories are particularily represented in the sub-database
- The diagnosis (regarding the nb. of records) seems hence less transparent



EDU & business unnafiliated to the financial and banking field (BSO, BSR)

- Higher rate of severe claims

Hypotheses :

- Those type of organisations are "inferior" when it comes to doing a diagnosis
- They do not wish to or do not know how to quantify the loss related to the lost data

The issue of claim management

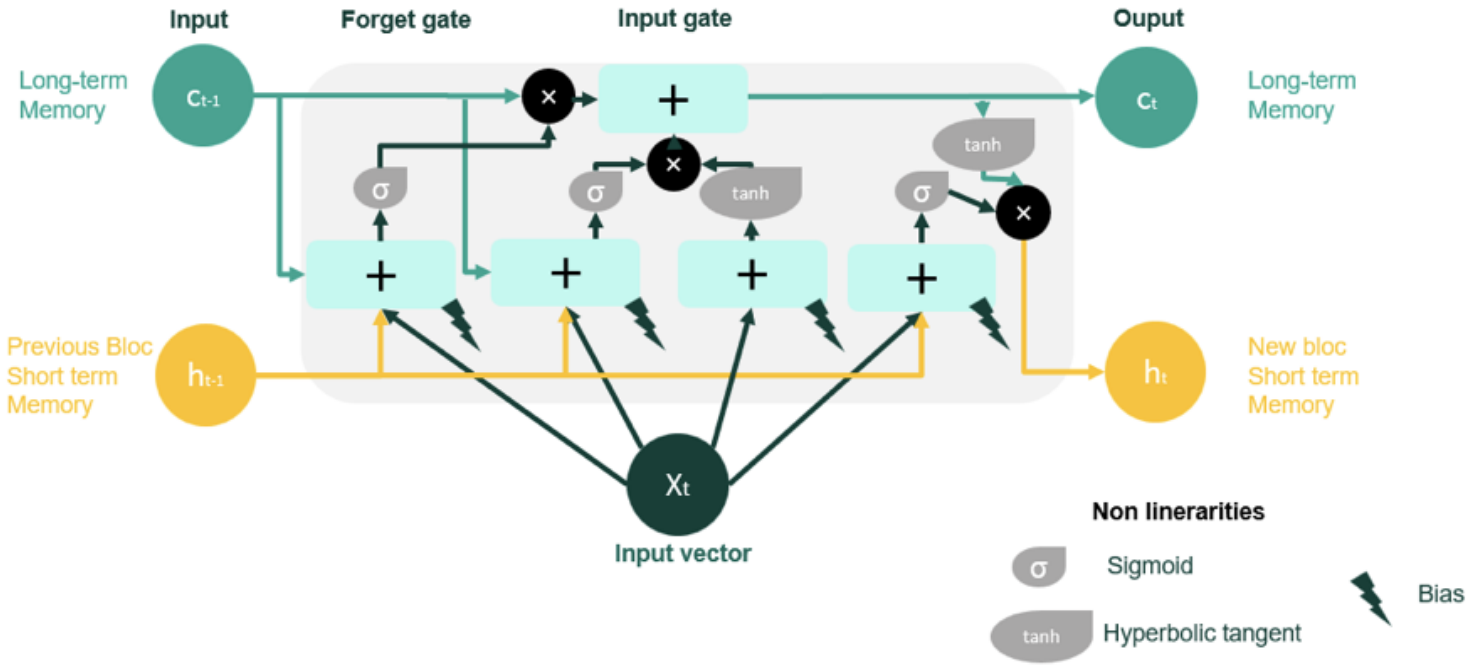
- Cyber insurance has a component of **victim assistance**

- Despite the assistance business not being new, the insurer lacks expertise in the cyber component

- A use of the presented methods : **Help detecting the claims that need a particular type of answer, to improve the way they are handled** (and to minimize their detrimental consequences)

- **Input** : Claim reports, preliminary expertise
- **Output** : diagnosis and assistance recommendations
- **Extension** : Follow-up of the evolution with time of the claims

Recurrent Neural Network (RNN) for text analysis



- Example of the LSTM
- Other reference on connex issue : Cohen-Sabban,I.,Lopez,O.,Mercuzot,Y. (2021) *Automaric analysis of insurance reports through deep neural networks to identify severe claims,* ***Annals of Actuarial Science.***

Perspective of evolution : the bayesian

Differences in the type of vision :

- **Frequentist approach** : we have $(X_1, \dots, X_n)$ of distribution $\mathbb{P}_{\theta_0}$, and we infer/estimate θ_0 using those only informations

- **Bayesian approach** : We suppose θ_0 to be a random variable, of given distribution *a priori* π, and we observe $(X_1, \dots, X_n)$ of distribution, given $\theta_0 = t$, $\mathbb{P}_t$

- The *a priori* approach implies a preliminary insight/diagnosis
- Question : how can we transform this insight mathematically speaking ?

Thank you

h.rapior@detralytics.eu

https://detralytics.com/detra-notes/