Deep Learning and Machine Learning in Industry, 3 applications

Dr Stefano Bromuri

Outline

What are Machine Learning and Deep learning

- Machine Learning and Machine Learning Process
- Feed Forward Networks
- LSTMs/CNNs
- NTMs and Encoders

What is representation learning

- Word2Vec
- Node2vec

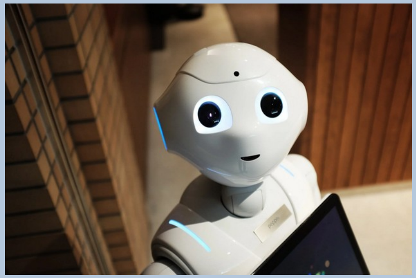
Motivating use cases

- -Emotion recognition (the VERA project, Customer Management)
- -HSCodes Prediction (Neural Machine Translation, Logistics)
- -Representation Learning for Rule Learning (Customer Activation)

AI/Machine Learning/Deep Learning

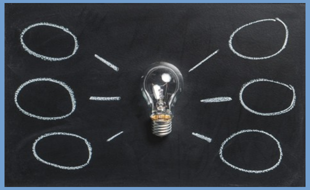
Artificial Intelligence

Techniques mimicking human behaviour



Machine Learning

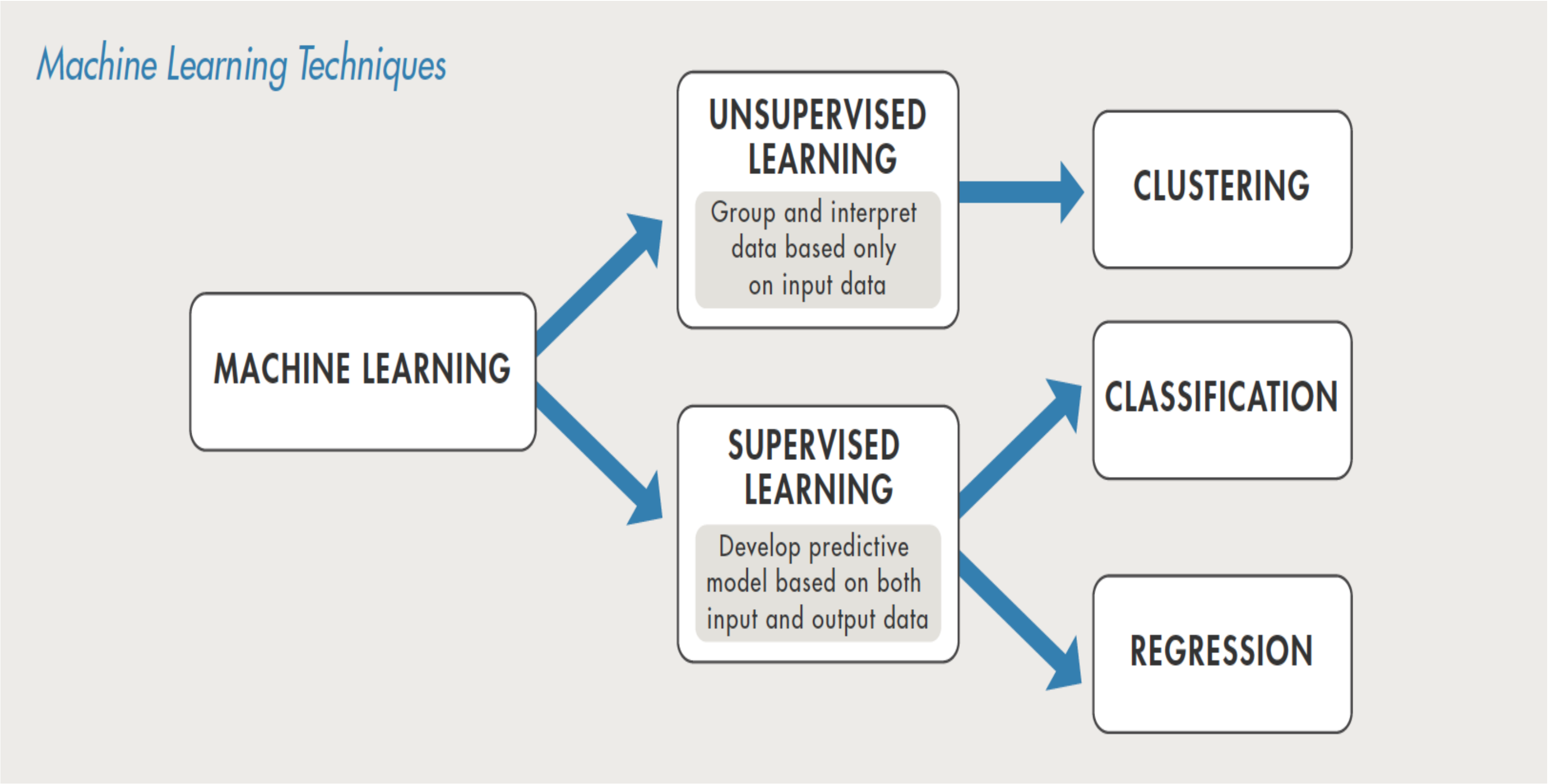
Learn from data after specifying the features



Deep Learning

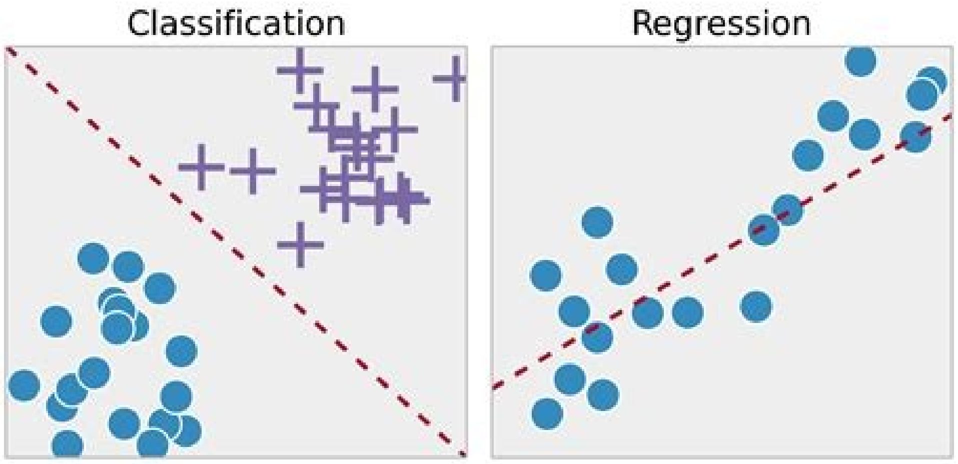
Learn the features while learning from data

What is Machine Learning?

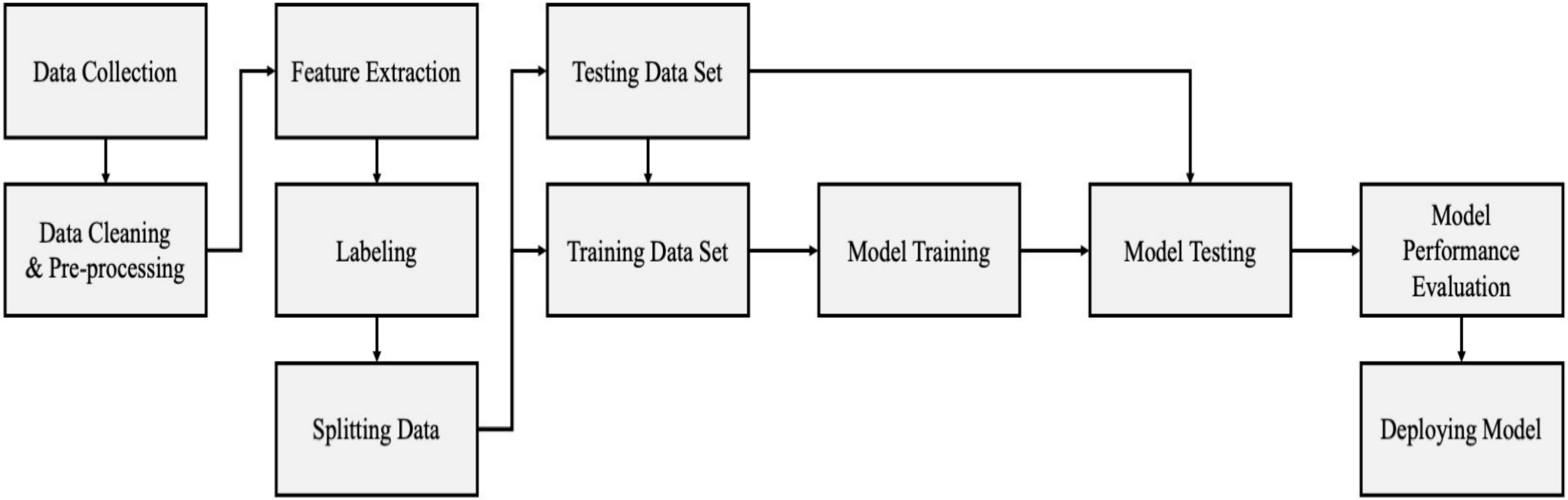


What is Machine Learning?

Machine learning trains mathematical functions to distinguish cloud of points in classes (classification or to find a trend that fits the cloud of points (regression).



Machine Learning Process

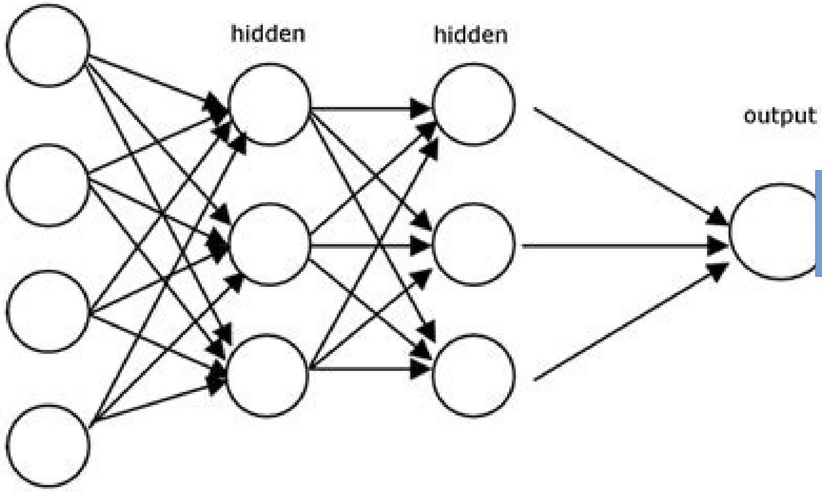


What is Deep Learning?
Feedforward network

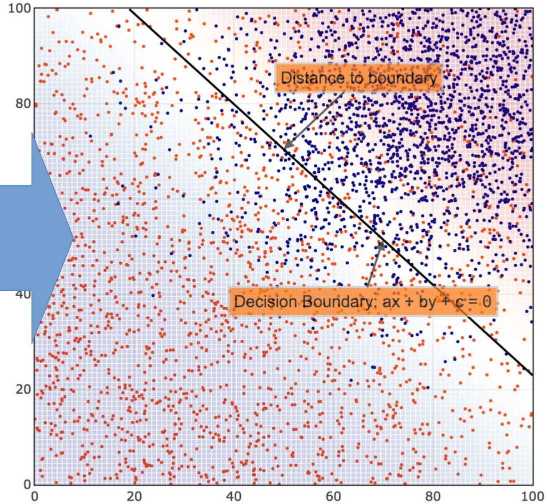
Flat Data

	Full Name	Hire Date	Location	State	Termination Date	Employment Type	Year	Base Salary	Bonus	Overtime	Commission
2	Alison Johnson	5/13/2000	Boston	MA		Full-Time	2000	$89,000	$6,230	$0	
3	Corine M. 'Henderson	8/7/2000	Boston	MA		Full-Time	2000	$72,000	$7,920	$2,880	
4	Julia Hegwood	8/24/2000	Bos"to"n	MA		Full-Time	2000	$45,000	$3,150	$0	$14,000
5	Jeremiah De Grazia	9/8/2000	Boston	MA		Full-Time	2000	$58,000	$4,640	$1,740	
6	Willow Nevandro	10/18/2000	Boston	MA		Full-Time	2000	$34,000	$2,380	$1,020	
7	Martie Elmasian	12/7/2000	Boston	MA		Full-Time	2000	$54,000	$6,480	$2,160	
8	Alison Johnson	5/13/2000	Boston	MA		Full-Time	2001	$92,000	$7,360	$0	
9	Corine M. Henderson	8/7/2000	Boston	MA		Full-Time	2001	$75,000	$5,250	$3,750	
10	Julia Hegwood	8/24/2000	Boston	MA		Full-Time	2001	$48,000	$2,880	$0	$44,000
11	Jeremiah De Grazia	9/8/2000	Boston	MA		Full-Time	2001	$62,000	$3,100	$3,100	
12	Willow Nevandro	10/18/2000	Boston	MA		Full-Time	2001	$36,000	$2,520	$1,080	
13	Martie Elmasian	12/7/2000	Boston	MA		Full-Time	2001	$60,000	$4,200	$2,400	
14	Kelly Queen	1/13/2001	Los Angeles	CA		Full-Time	2001	$43,000	$5,160	$0	$42,000
15	Cristhian Roth	4/2/2001	Los Angeles	CA		Full-Time	2001	$54,000	$4,320	$0	
16	Joeanne Melendez	5/4/2001	Los Angeles	CA		Full-Time	2001	$32,000	$2,560	$0	
17	Brian M Stucki	5/10/2001	Chicago	IL		Full-Time	2001	$43,000	$3,440	$0	$22,000
18	Cindy Summerville	6/23/2001	Chicago	IL		Full-Time	2001	$63,000	$3,780	$0	
19	Tony Merrick	9/18/2001	Chicago	IL		Full-Time	2001	$105,000	$9,450	$3,150	
20	Bryan Anderson	10/6/2001	Chicago	IL		Full-Time	2001	$66,000	$3,960	$3,300	

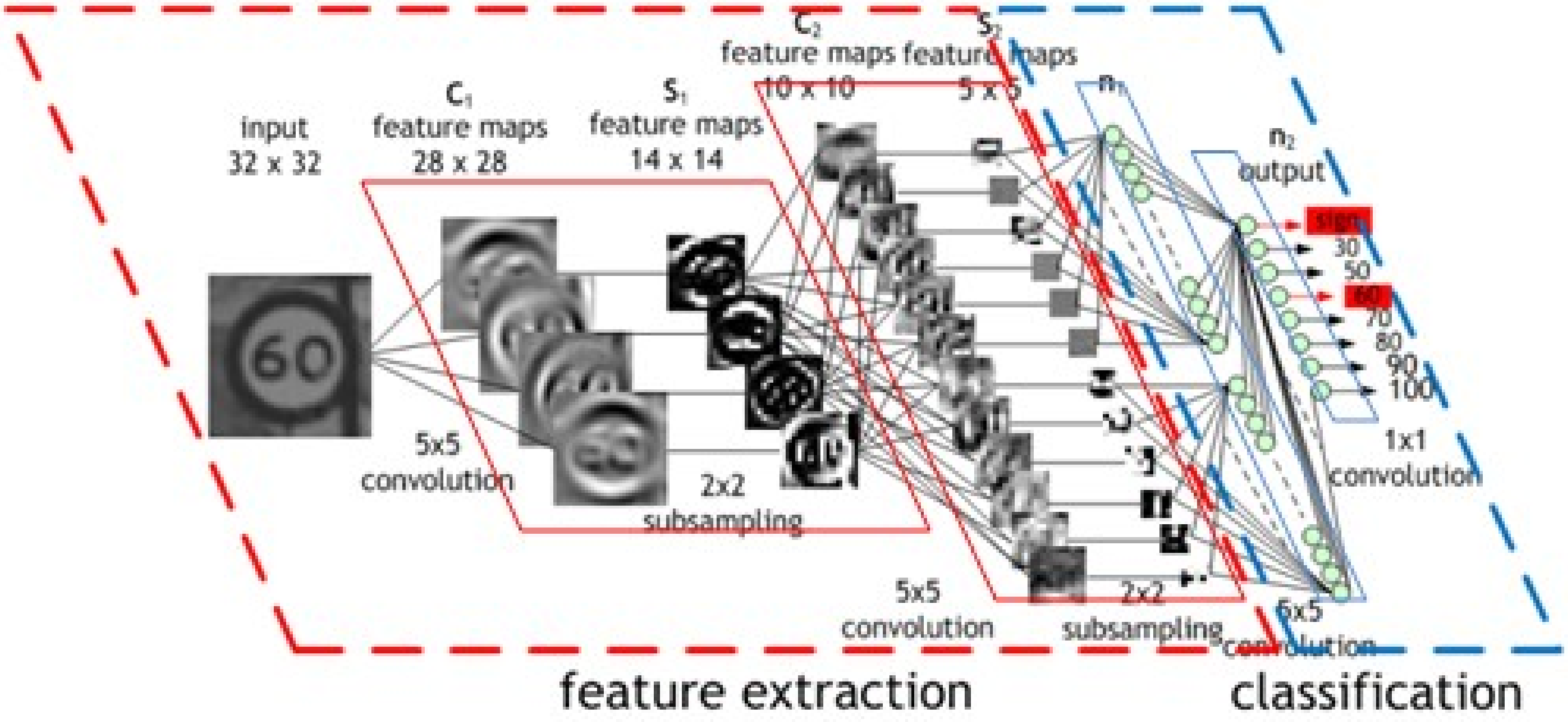
“Neurons”



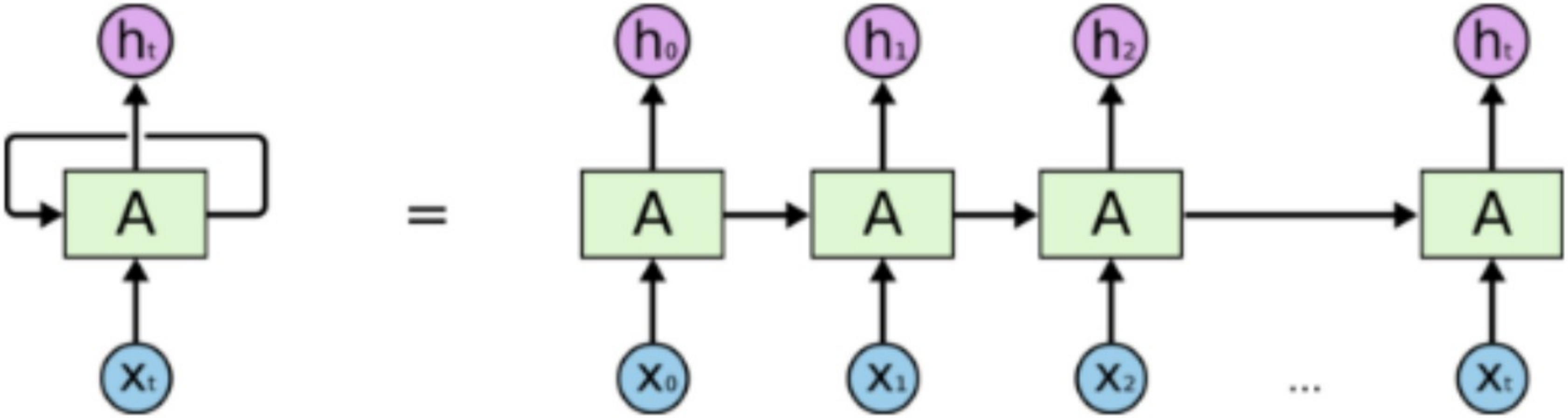
Equation of classificati
boundary



What is Deep Learning?
Convolutional Network (for images)

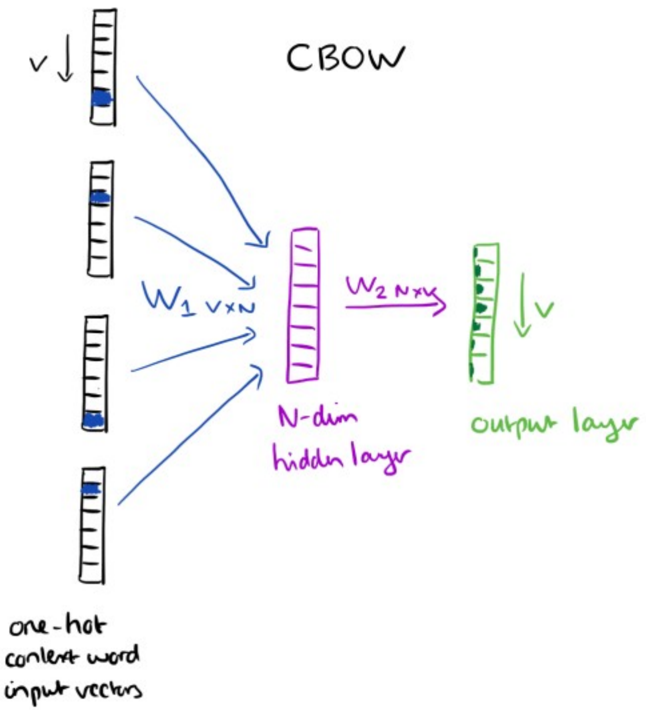


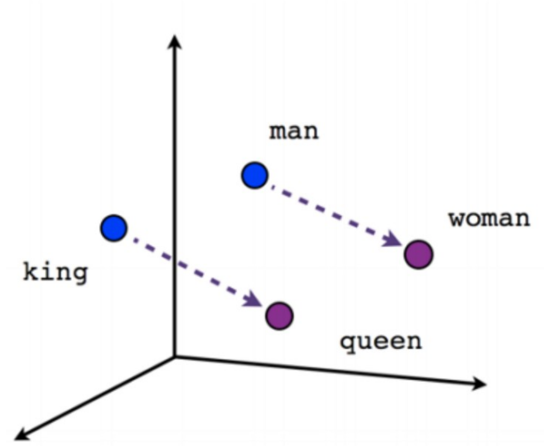
What is Deep Learning?
LSTM Network (for sequences)



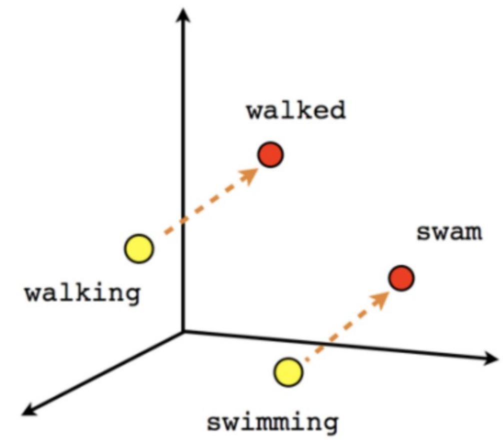
An unrolled recurrent neural network.

Representation Learning, Word2Vec



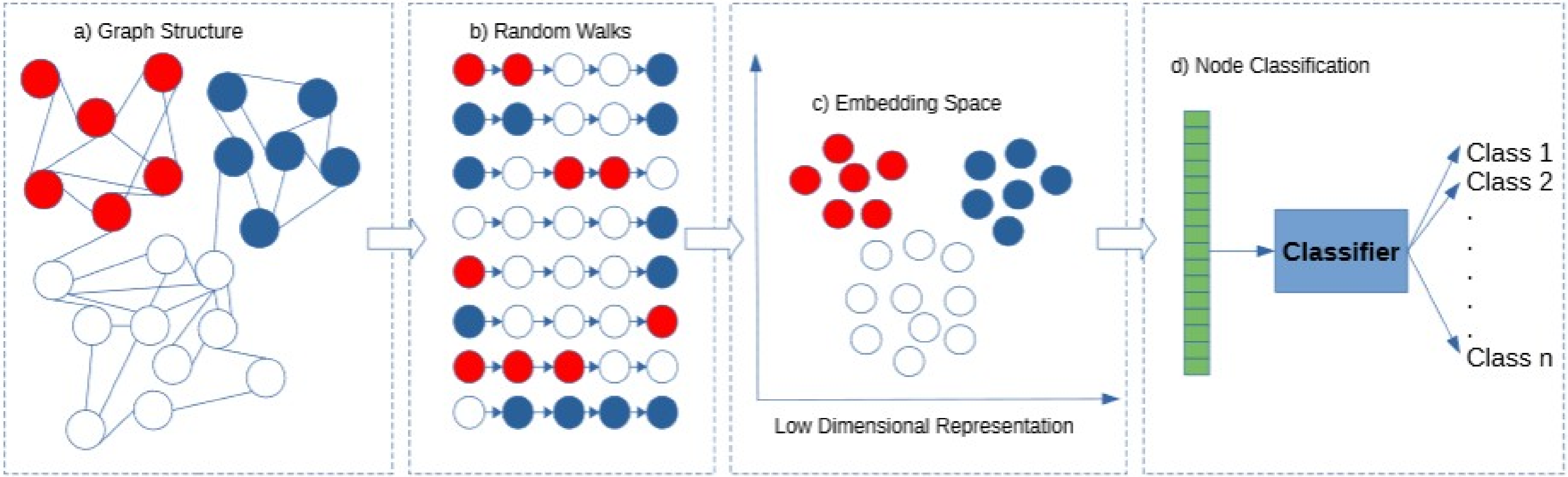


Male-Female



Verb tense

Representation Learning, Node2Vec



Motivating Use Cases

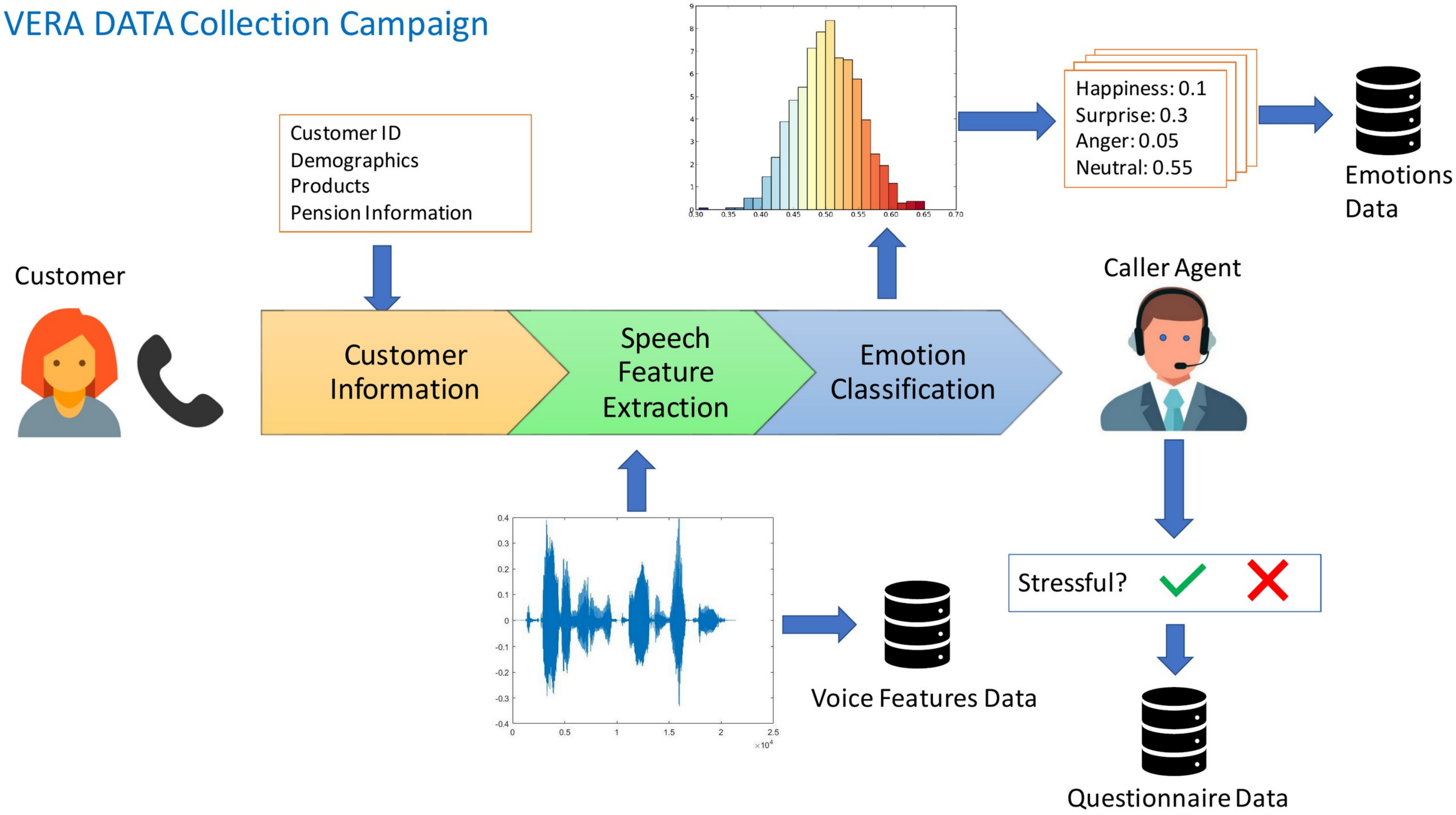
3 Applications from Industry:

Emotion recognition (the VERA project, Customer Management)

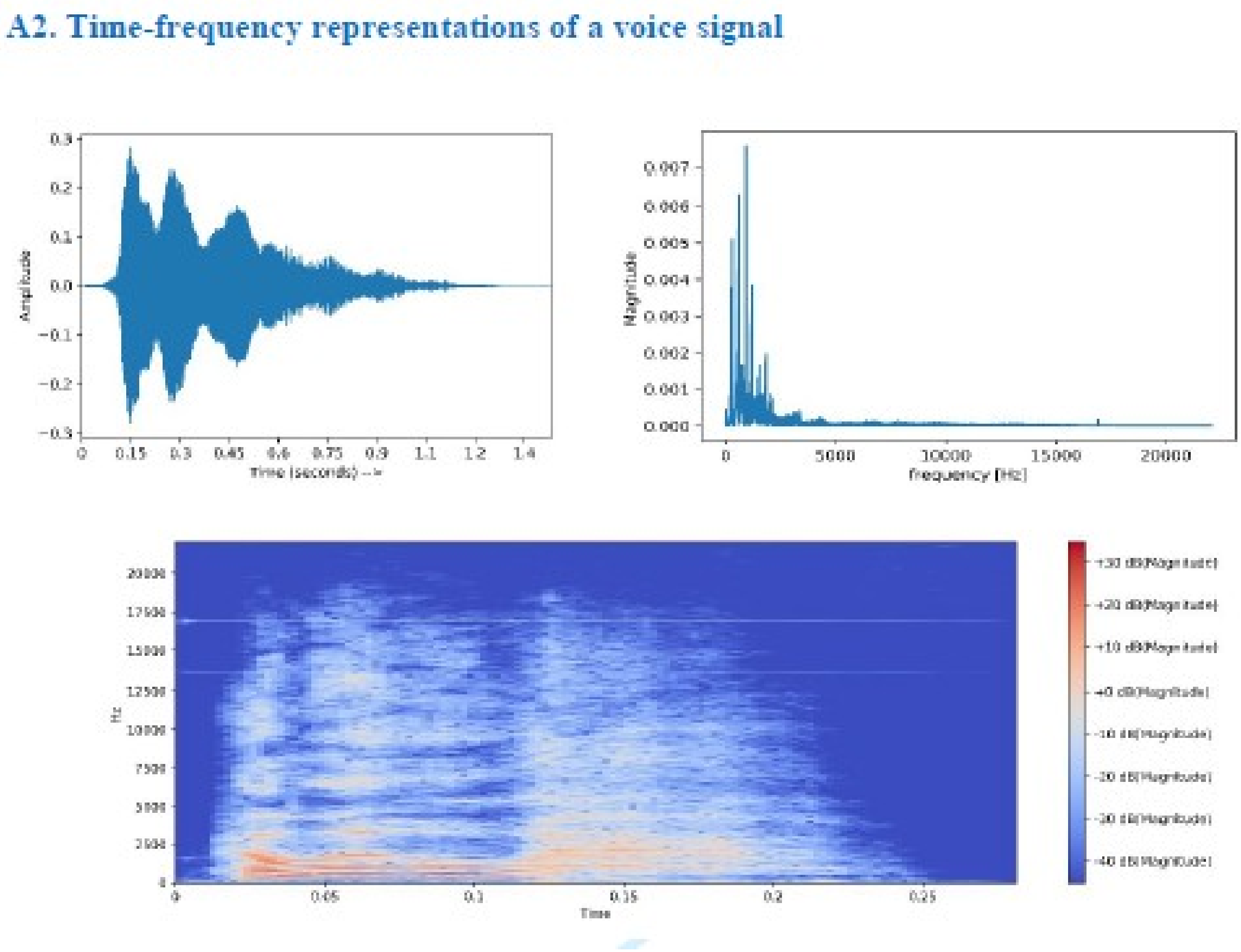
-HSCodes Prediction (Neural Machine Translation, Logistics)

-Representation Learning for Rule Learning (Customer Activation)

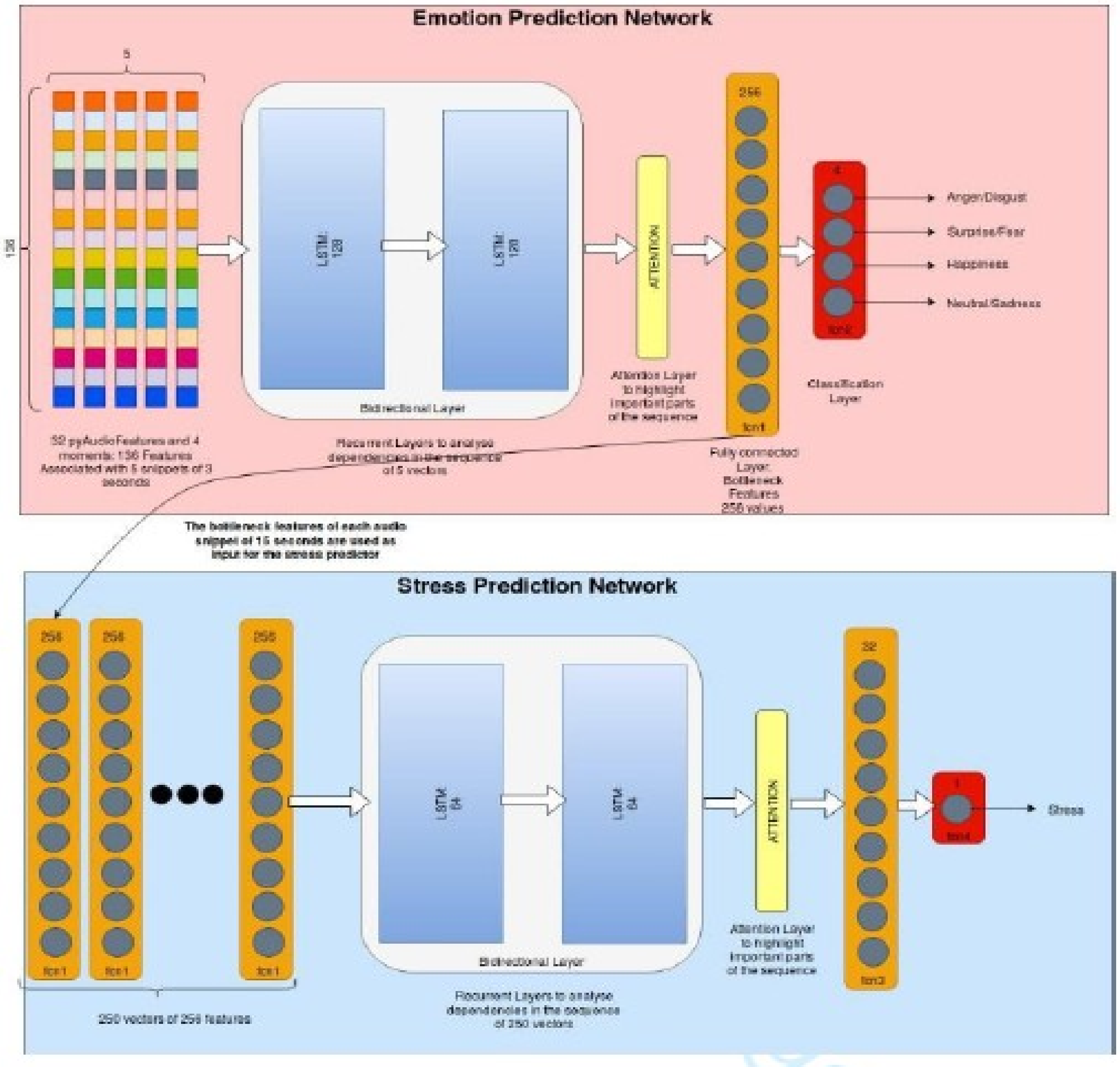
Motivating Use Cases: VERA



VERA Voice Signal



VERA Architecture



Results: Emotion and Stress Prediction

	1DLA			LA			BIL			Frequency Classifier		
	Prec.	*Recall*	*f1*	*Prec.*	*Recall*	*f1*	*Prec.*	*Recall*	*f1*	*Prec.*	*Recall*	*f1*
Anger / Disgust	0.21 (0.04)	0.12 (0.04)	0.17 (0.04)	0.21 (0.04)	0.17 (0.03)	0.19 (0.03)	0.27 (0.05)	0.18 (0.03)	0.22 (0.03)	0.08 (0.03)	0.09 (0.03)	0.08 (0.03)
Fear / Surprise	0.23 (0.04)	0.25 (0.04)	0.19 (0.04)	0.19 (0.03)	0.19 (0.03)	0.19 (0.03)	0.21 (0.03)	0.21 (0.03)	0.21 (0.03)	0.12 (0.03)	0.15 (0.03)	0.13 (0.03)
Happiness	0.25 (0.04)	0.07 (0.03)	0.1 (0.03)	0.24 (0.04)	0.26 (0.05)	0.27 (0.05)	0.28 (0.05)	0.17 (0.04)	0.21 (0.04)	0.1 (0.03)	0.09 (0.03)	0.09 (0.03)
Sadness / Neutral	0.78 (0.01)	0.88 (0.01)	0.83 (0.01)	0.8 (0.01)	0.81 (0.01)	0.81 (0.01)	0.8 (0.01)	0.86 (0.01)	0.83 (0.01)	0.73 (0.01)	0.68 (0.01)	0.71 (0.01)
Micro Avg.	0.67 (0.01)	0.67 (0.01)	0.67 (0.01)	0.68 (0.01)	0.68 (0.01)	0.68 (0.01)	0.68 (0.01)	0.68 (0.01)	0.68 (0.01)	0.53 (0.01)	0.53 (0.01)	0.53 (0.01)

Note. The upper statistic represents the mean and the lower statistic in-between brackets represents the 95%-confidence interval ($p < .05$).

Table 4. Stress prediction network performance.

Stress network predictor	**Precision**	**Recall**	**f1-score**
No-stress	0.86 (0.03)	0.84 (0.03)	0.85 (0.03)
Stress	0.68 (0.06)	0.7 (0.06)	0.69 (0.06)
Micro-avg	0.8 (0.03)	0.8 (0.03)	0.8 (0.03)

Note. The upper statistic represents the mean and the lower statistic in-between brackets represents the 95%-confidence interval ($p<.05$).

VERA: Ethical Problems

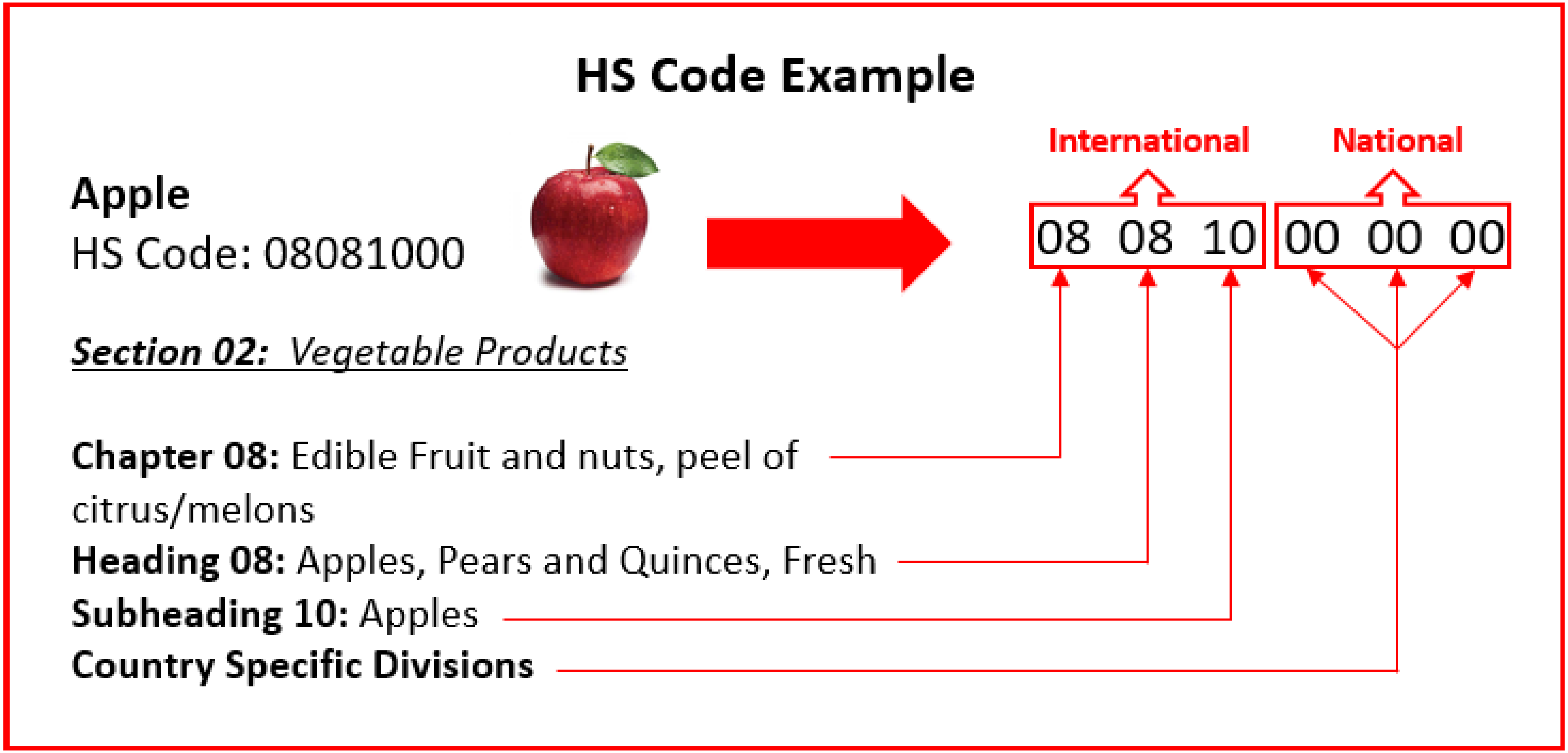
- Usage of emotions patterns to scam people into doing something

- Usage of emotion patterns to lure information

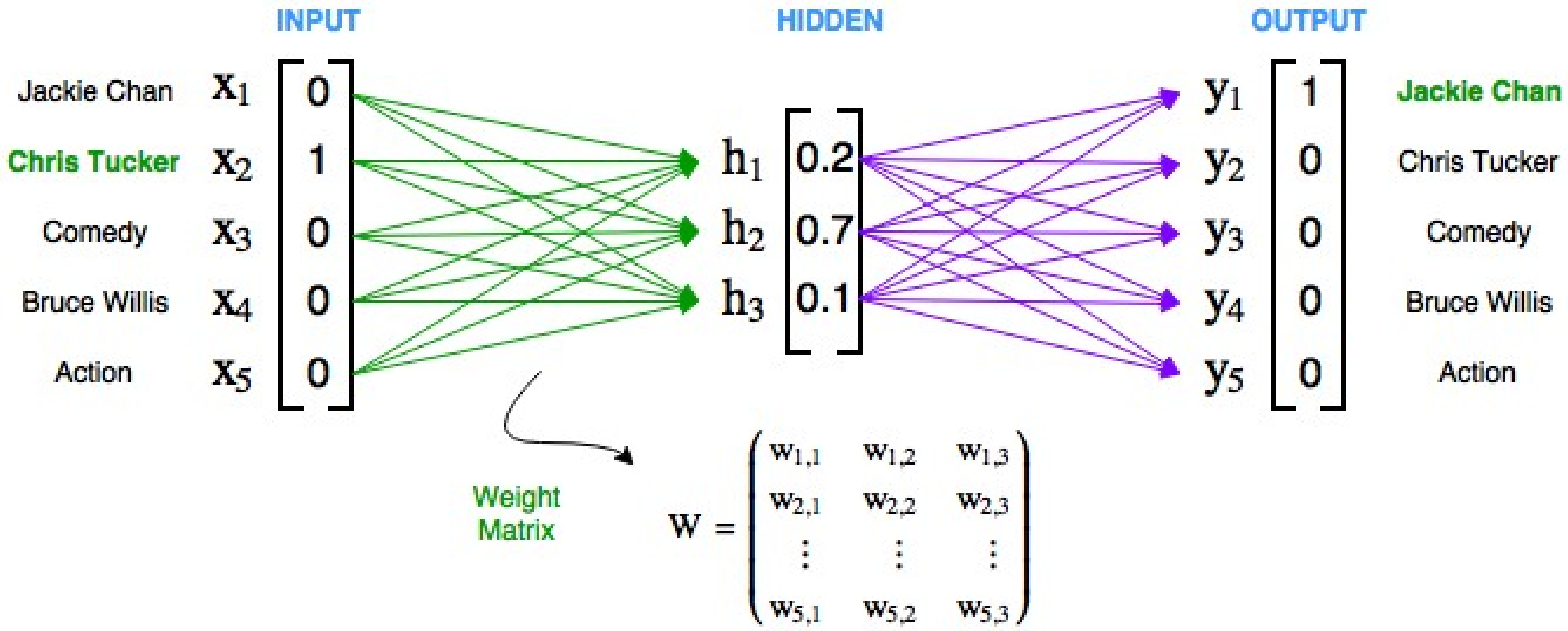
- Usage of emotion patterns to fire personnel

- Profiling and tracking people based on emotion responses

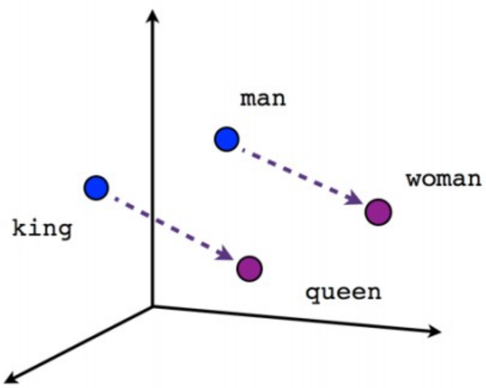
Motivating Use Cases: HS Codes



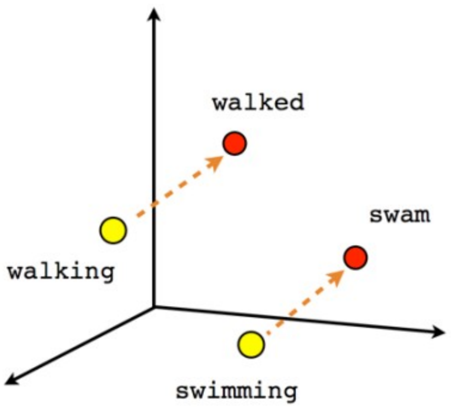
Component 1: Word Embeddings



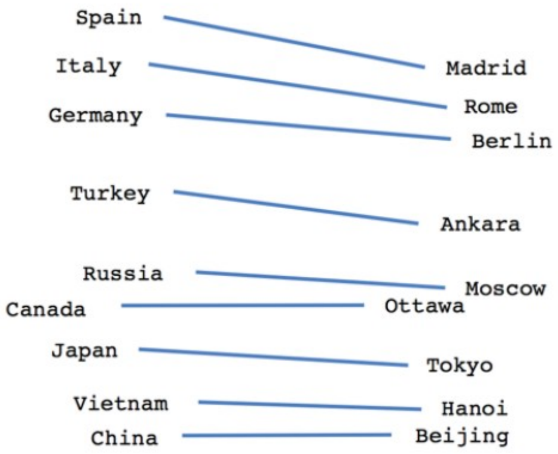
Component 1: Word Embeddings



Male-Female

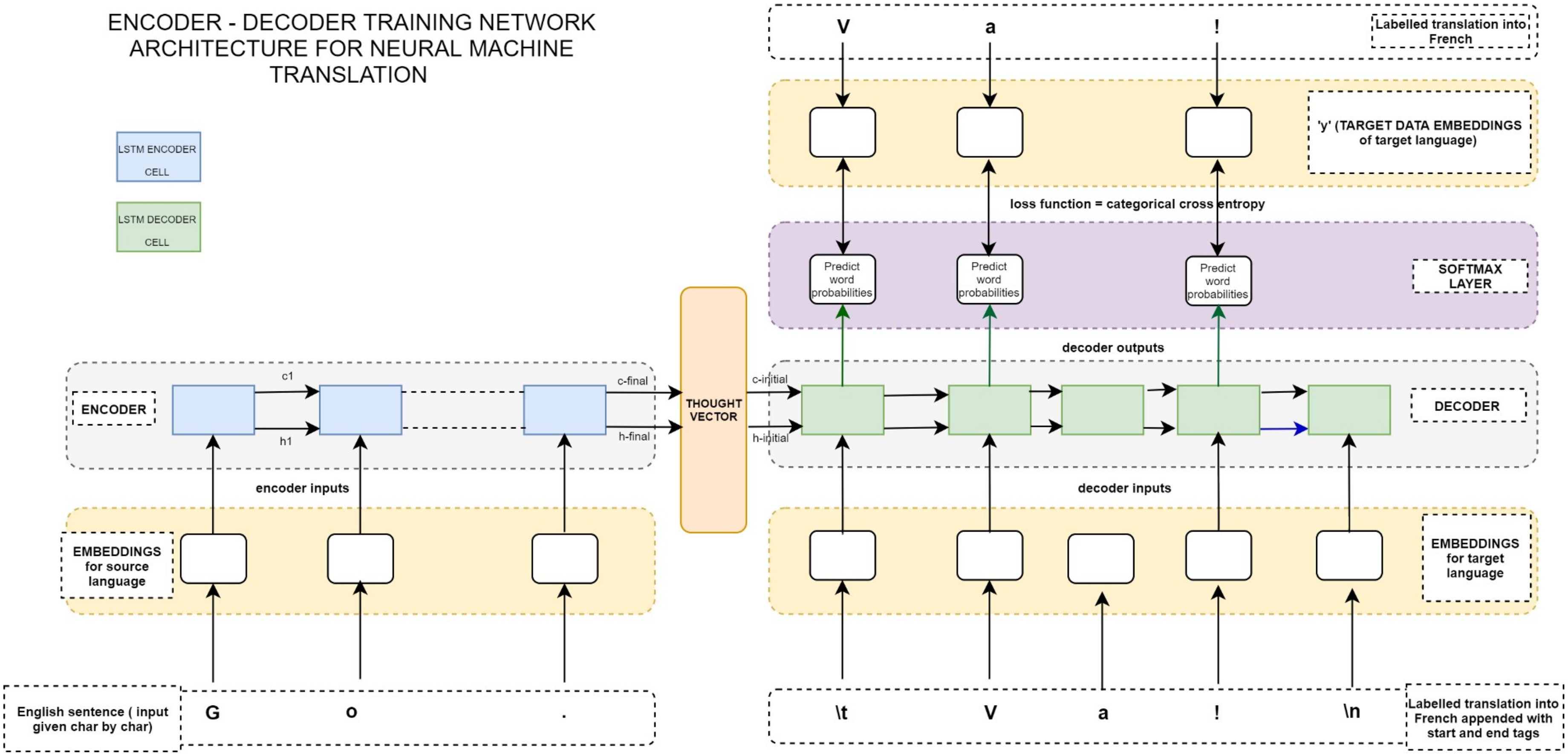


Verb tense

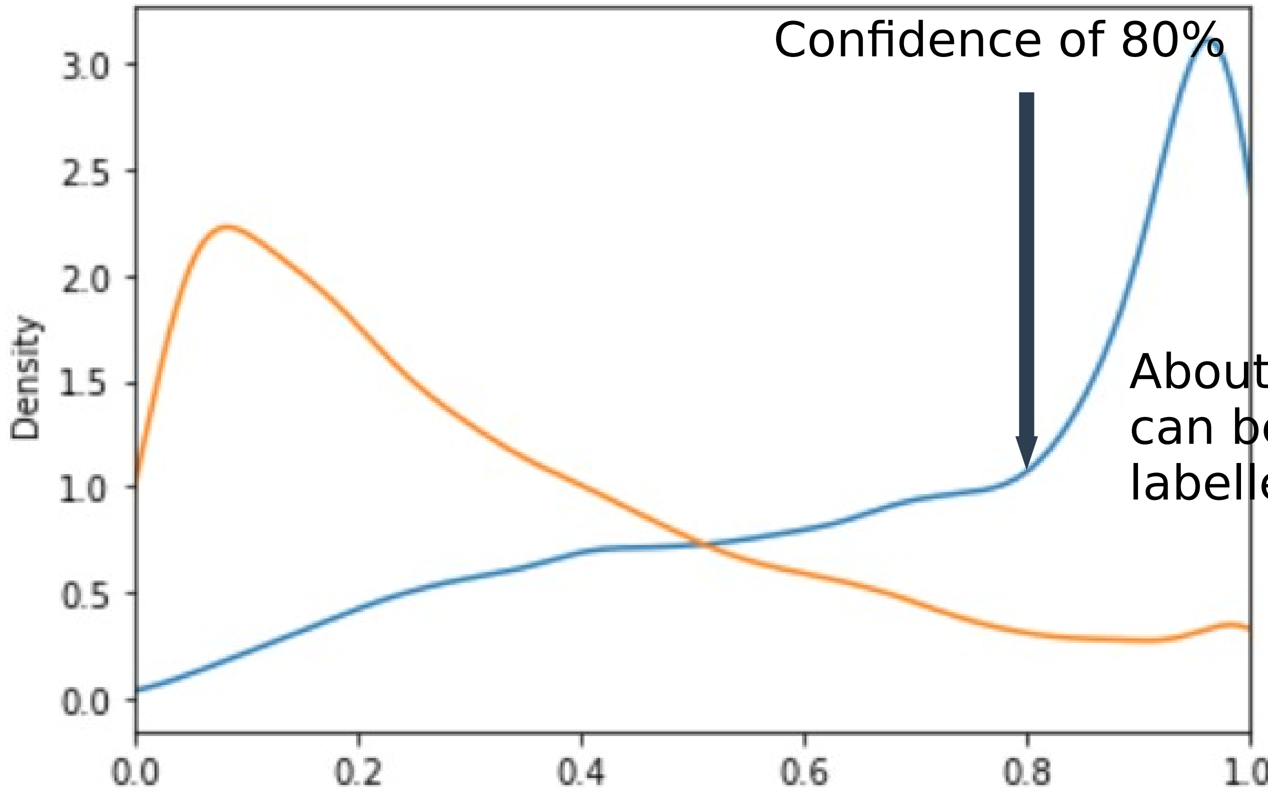


Country-Capital

Component 2: Neural Machine Translator



Quality Curves



About 16% of the data can be automatically labelled.

Results

- **The NMT succeeds at classifying about 16% of the data with a 80% confidence.**
- **This result is going to be replicated through 11 states (the biggest ones being Germany and UK)**
- **Any percentage point of saving is tens of thousands of euros.**

HS Codes: Ethical Problems

The main issue with **automating** everything is that less people will be hired to perform the job that the robot is taking care of.

Of course the same people could be used for something more **meaningful** if the process is particularly repetitive.

Profiling is also a problem, although less accentuated in this particular case, given that a code has to be provided by **law**.

Motivating Use Case, Segmenting Customers for Customer Activation

Pension funds send news letters to:

Advise the customer concerning their pension

Changes in the law

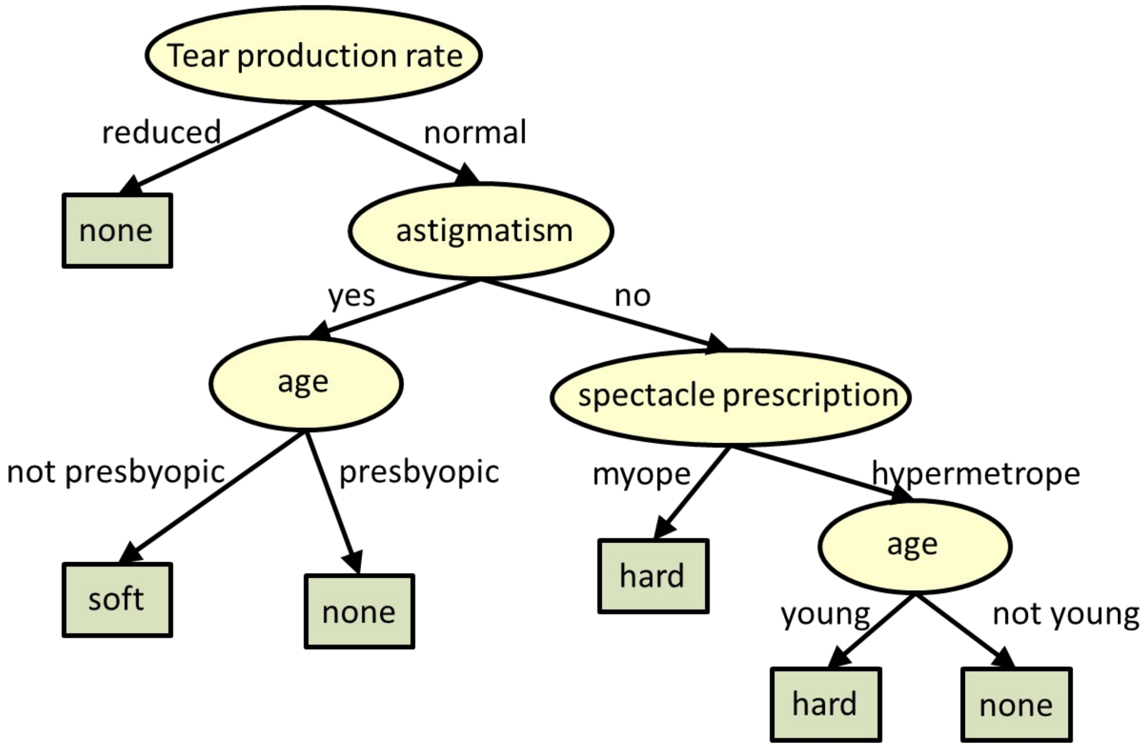
Notify a new available service

Notify changes in the taxation

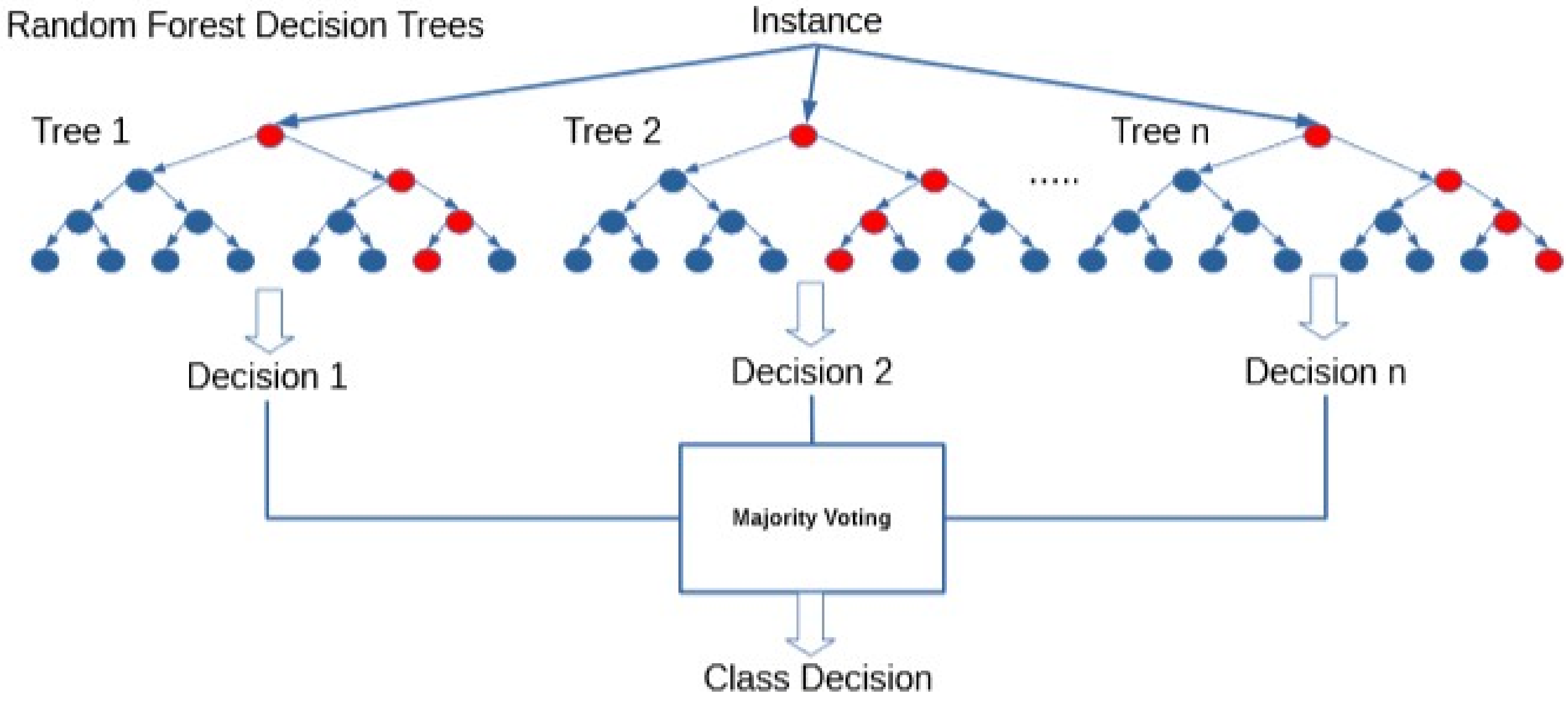
The pension fund really needs that as many customers as possible check the information, but the budget is limited!
→ **Machine Learning to segment the customers!**

The problem: Machine learning algorithms are black boxes (?)

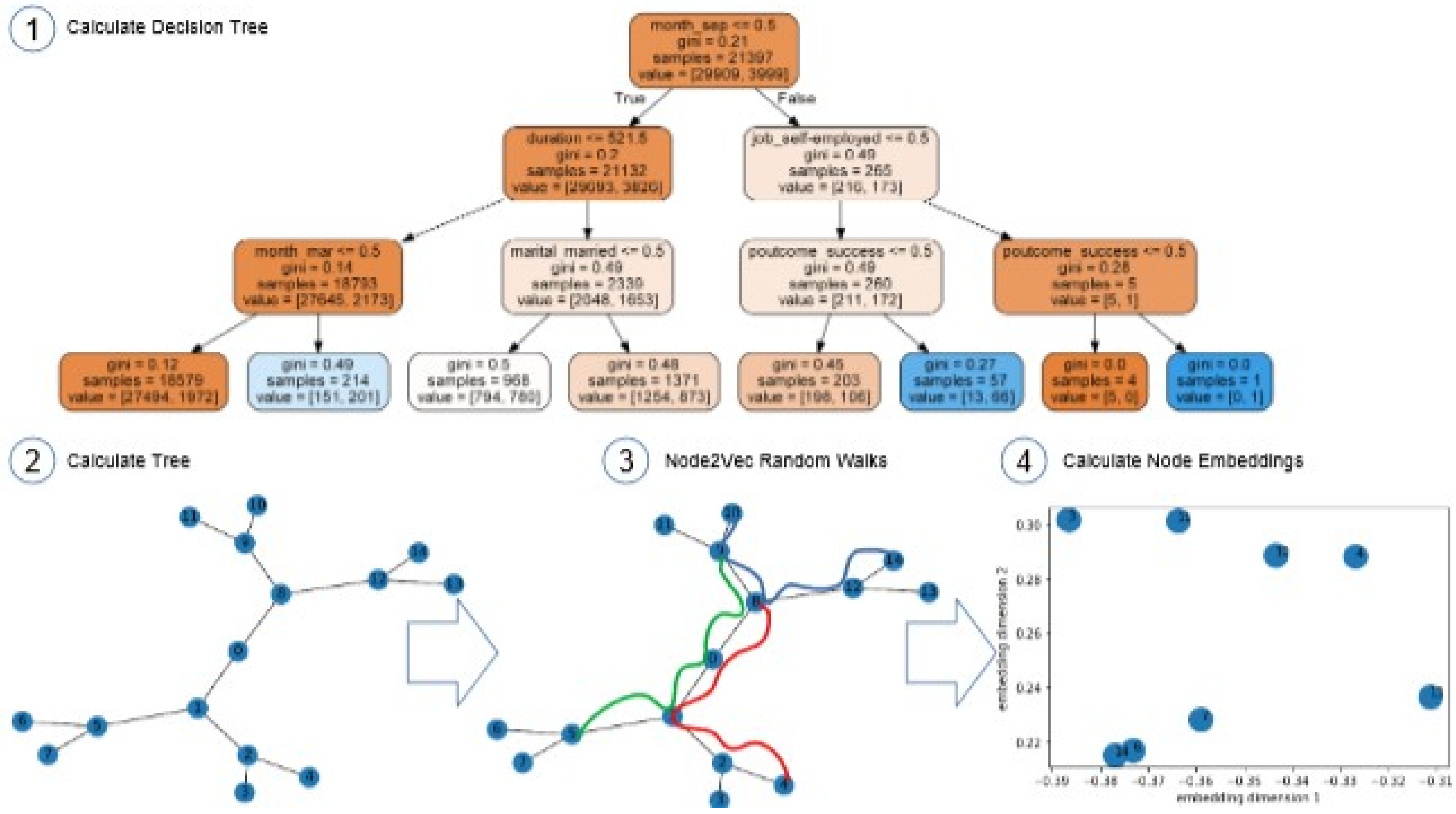
Not always, Decision trees are not black box. They define rules.



Motivating Use Cases: Random Forest Node Embeddings



Random Forest Node Embeddings Example

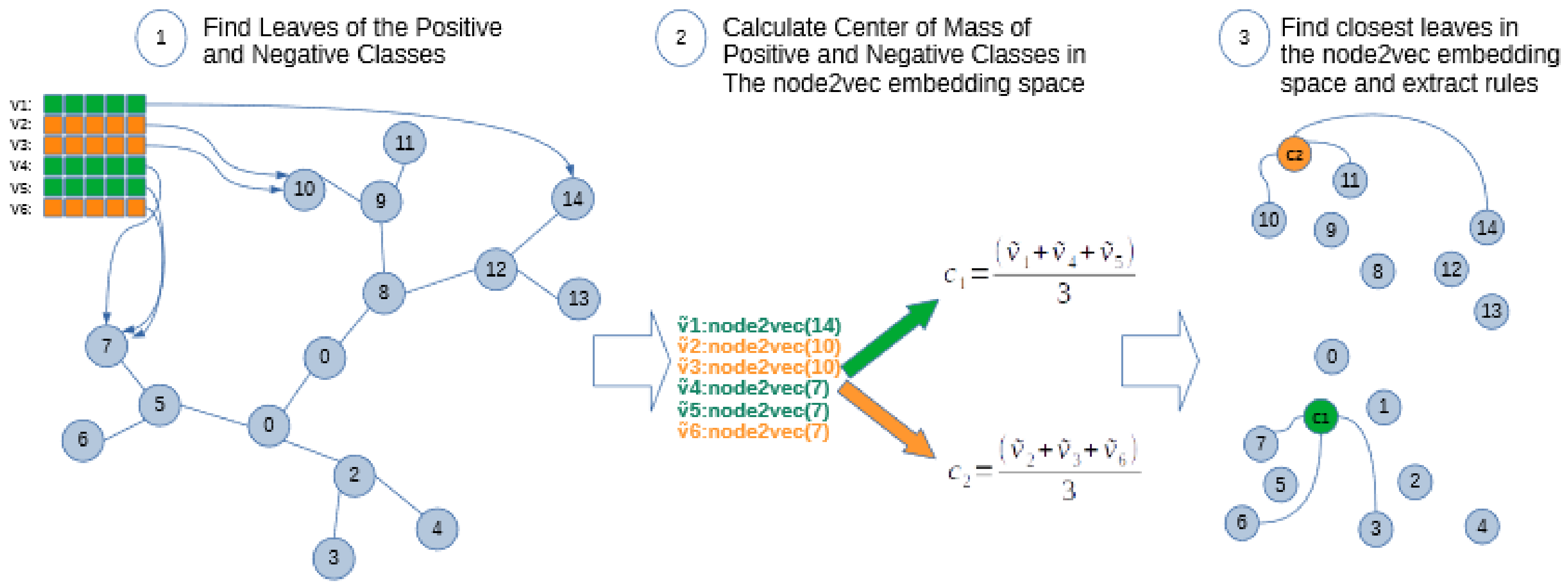


RFNE Some Results, classify

Method	D1	D2	D3	D4
KNN	0.849	0.533	0.566	0.931
Naive Bayes	0.0729	0.568	0.615	0.90
Logistic Regression	0.861	0.608	0.641	0.941
RandomForest	**0.922**	**0.657**	0.707	0.947
ExtraTrees	0.920	0.624	0.686	0.945
Random Forest Node Embeddings	0.907	**0.657**	**0.722**	**0.948**

It does not lose information wrt RandomForests,
so it is a plausible feature extraction

RFNE, Describe



RFNE, Describe

	log odds	stde	z	P> $\|z\|$	[0.025	0.975]	odds
Intercept	-4.0580	0.186	-21.846	0.000	-4.422	-3.694	0.017284
feature_0	0.2344	0.134	1.751	0.080	-0.028	0.497	1.264150
feature_1	0.2812	0.078	3.620	0.000	0.129	0.433	1.324719
feature_2	0.3986	0.067	5.967	0.000	0.268	0.530	1.489738
feature_3	0.7227	0.098	7.346	0.000	0.530	0.916	2.059988
feature_4	-0.2204	0.069	-3.201	0.001	-0.355	-0.085	0.802198
feature_5	0.0864	0.093	0.924	0.355	-0.097	0.270	1.090242
feature_6	-0.1506	0.108	-1.392	0.164	-0.362	0.061	0.860192
feature_7	-0.0289	0.076	-0.378	0.706	-0.179	0.121	0.971514
feature_8	-0.1441	0.098	-1.473	0.141	-0.336	0.048	0.865801
feature_9	0.6914	0.096	7.192	0.000	0.503	0.880	1.996509
feature_10	-0.2434	0.133	-1.824	0.068	-0.505	0.018	0.783958

$$
\begin{aligned}
rule_{d1} = & \text{pdays} < 9.5 \text{ and marital != 'single' and} \\
& \text{job} \in [\text{"admin.", "blue-collar", "entrepreneur", "housemaid"}] \text{ and} \\
& \text{age} < 61.5 \text{ and day} > 18.5
\end{aligned}
$$

Ethical Problems

- Segmenting Customers to hypertarget them with the perfect advertisement.

- Scamming customers very effectively and cheaply.

- Grouping Customers unfairly, without their consent.

Conclusion

- **AI offers many advantages**
  - Automation
  - Targeting population
  - Produce New insights
- **It comes with a number of negative aspects:**
  - People lose jobs
  - Scamming
  - Unfair use of data
  - Intellectual property

Questions ?