Brève introduction à l’apprentissage automatique

par **Julien Velcin**
Université Lumière Lyon 2, laboratoire ERIC
http://mediamining.univ-lyon2.fr/velcin

DIU Enseigner l'informatique au lycée
juin 2020

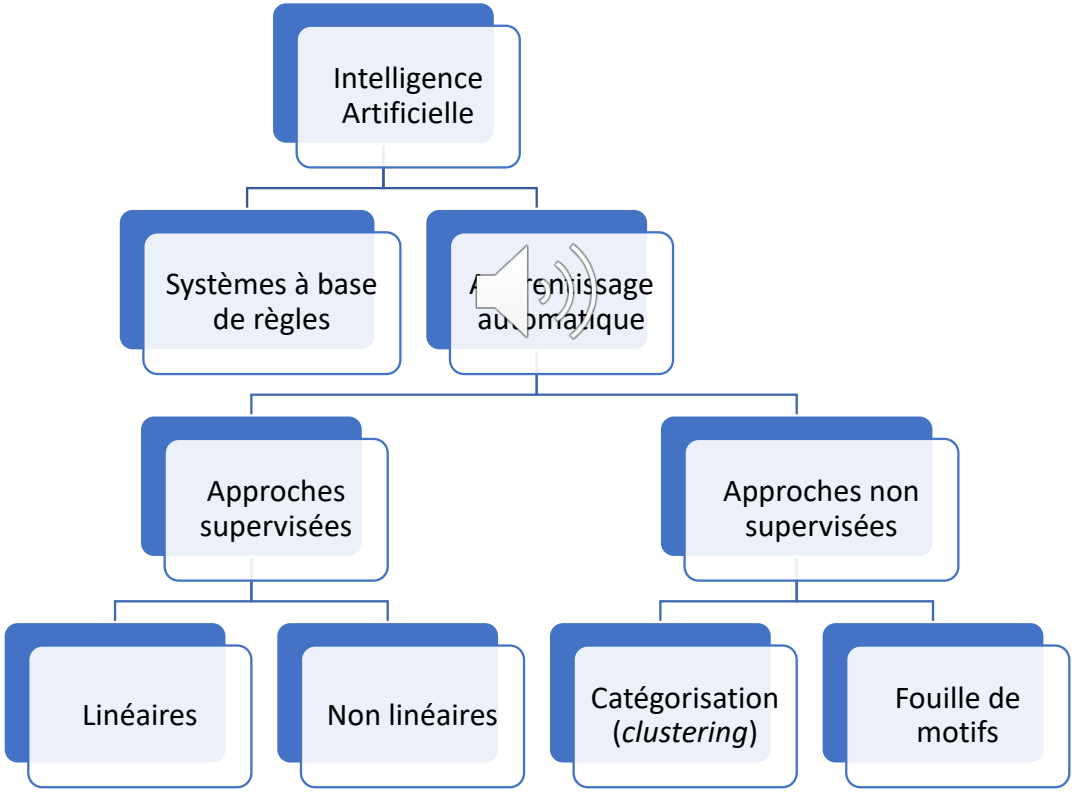
Plan du cours

- Objectifs de l’apprentissage automatique
- Exemples d’application
- Cas d’étude 1 : classification de documents
- Cas d’étude 2 : traduction automatique
- L’apprentissage automatique en pratique
- Références

Objectifs du « machine learning »

- L'objectif principal consiste à apprendre automatiquement à **généraliser** à partir d'exemples observés afin de pouvoir faire de l'inférence sur de **nouveaux exemples** jamais été observé auparavant (principe inductif)
- Différentes familles d'algorithmes :
  - apprentissage par coeur
  - apprentissage par cas
  - apprentissage par renforcement
  - classification non supervisée / catégorisation (*clustering*)
  - classification supervisée, régression
  - algorithmes génératifs

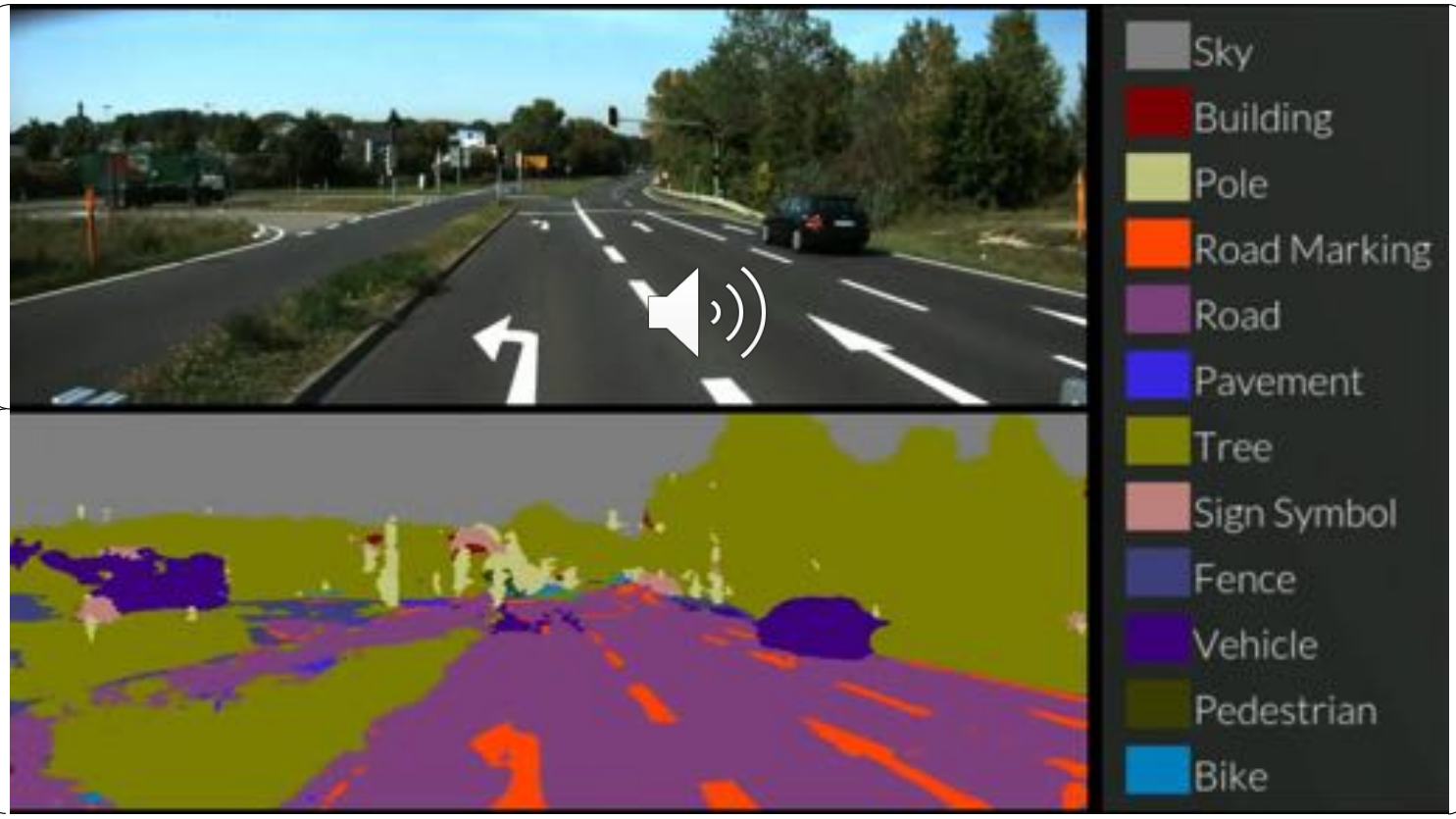
Positionnement vis-à-vis de l'IA



Exemple de la classification pour la reconnaissance d'objets dans des images

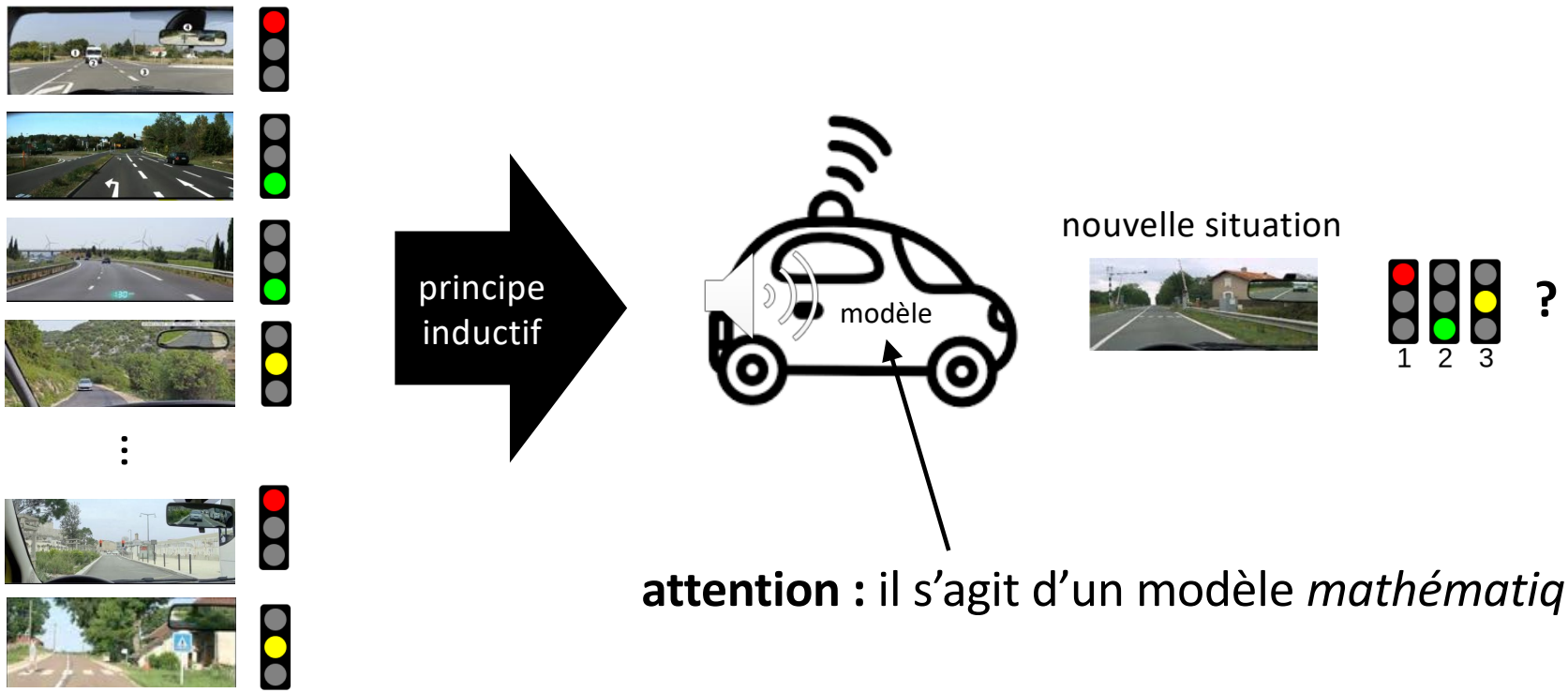
observations

ce qu'on cherche à prédire



classes

L'apprentissage automatique (ici, classification)



attention : il s'agit d'un modèle *mathématique !*

Exemples d’application

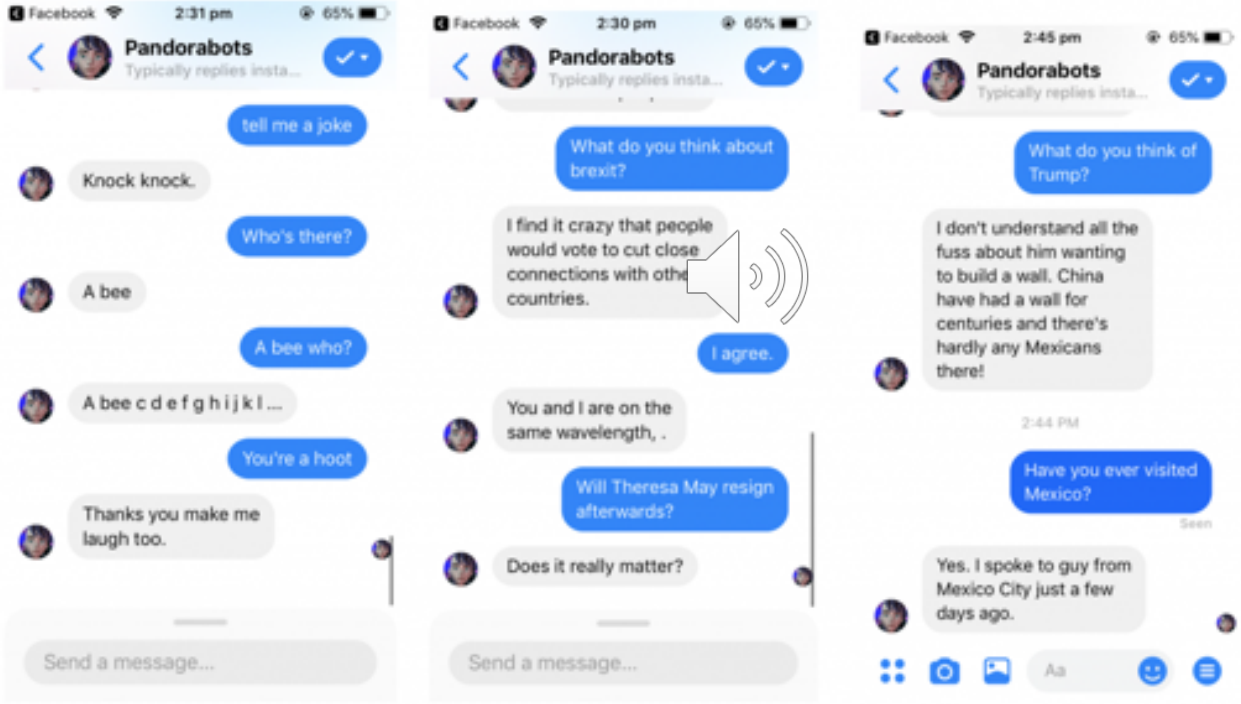
Julien Velcin

DIU EIL, 2019-2020

Jeux

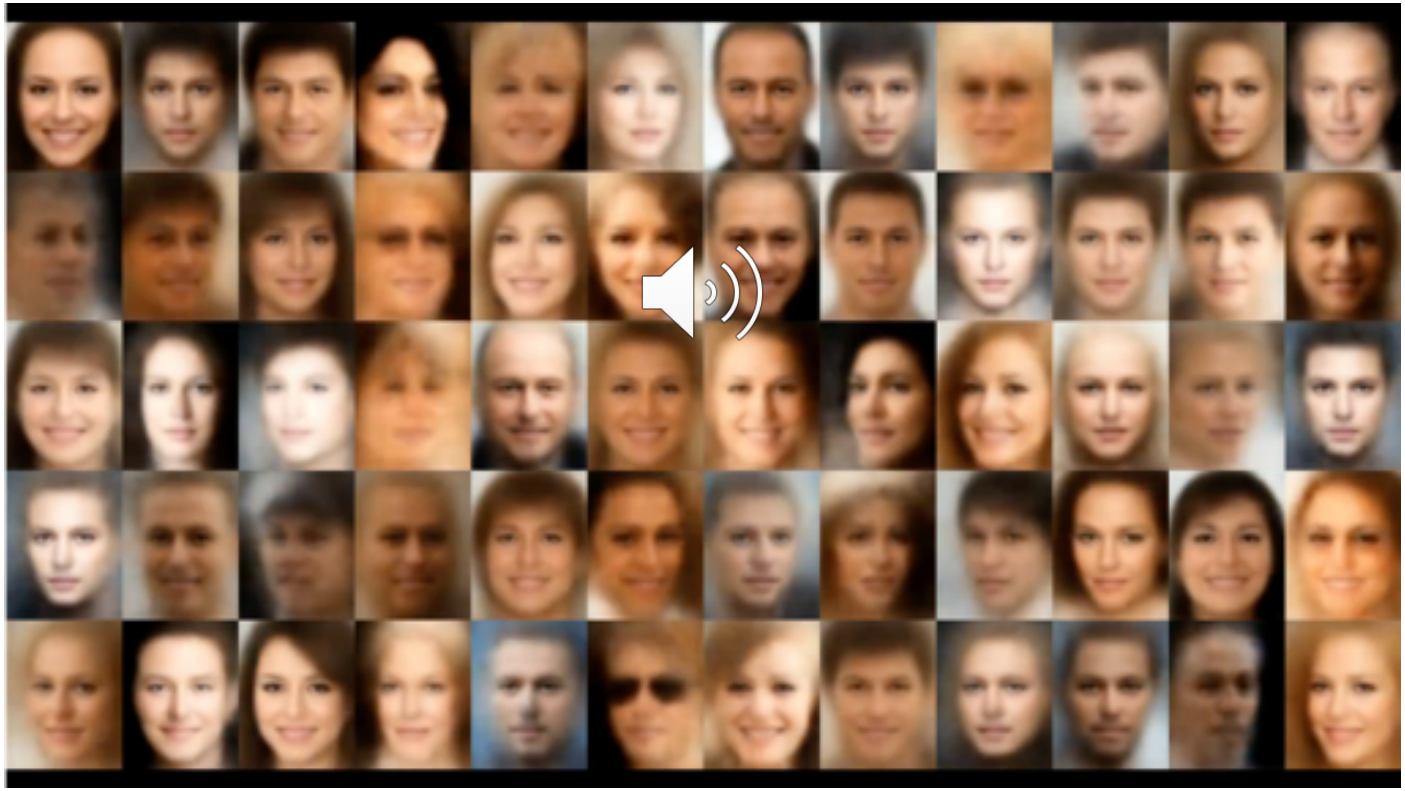
Apprendre à jouer
aux jeux vidéos

Robots conversationnels



https://www.pandorabots.com/mitsuku/

Génération de visages



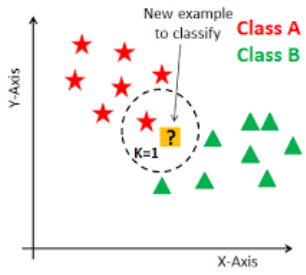
Cas d’étude 1 : classification de documents

Julien Velcin

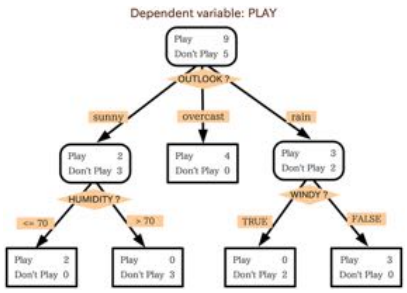
DIU EIL, 2019-2020

Classer des documents

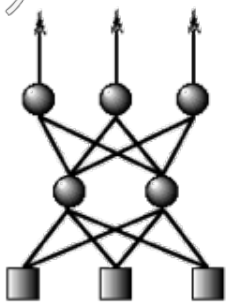
- Par exemple détecter la **polarité** d'un tweet ou classer un article dans une **thématique** déterminée (ex. sport ou économie)
- De (très) nombreux algorithmes existent :



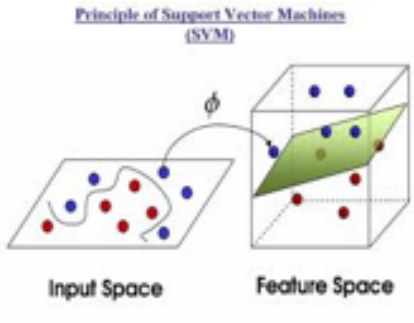
KNN



arbre de décision



réseaux de neurones artificiels



SVM

Procédure de traitement des données

termes décrivant les documents (vocabulaire)
par ex. T1 = « covid-19 », T2 = « football » et T3 = « masques »

prétraitement

	T1	T2	T3	T4	T5	T6	T7	T8	
Doc1	[illegible]	0	4	3	0	1	0	2	A
Doc2	0	2	4	0	2	3	0	0	B
Doc3	4	0	1	3	0	1	0	1	A
Doc4	0	1	0	2	0	0	1	0	B
Doc5	0	0	2	0	0	4	0	0	C
Doc6	1	1	0	2	0	1	1	3	A
Doc7	2	1	3	4	0	2	0	2	A

classes
par ex. A = « covid-19 »
B = « sport »
C = « plongée »

principe inductif

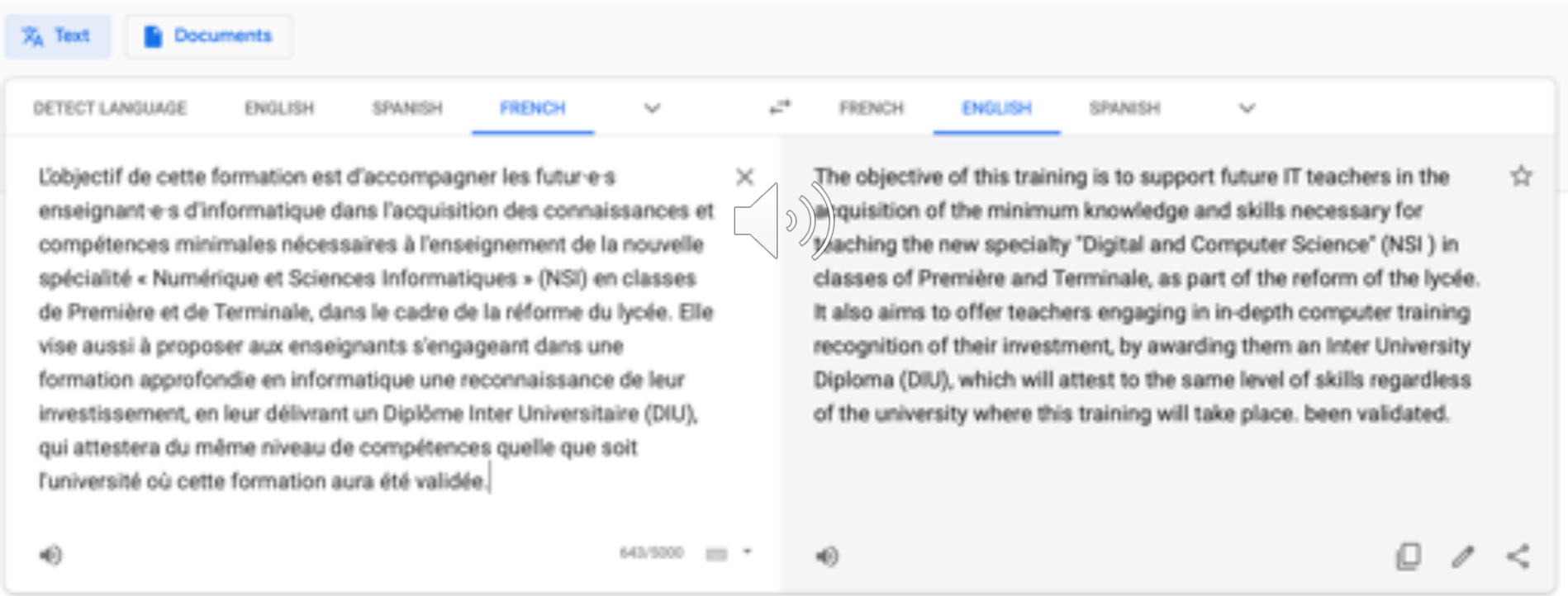
modèle pour la classification
d'un **nouveau** document Doc_n

Cas d’étude 2 : traduction automatique

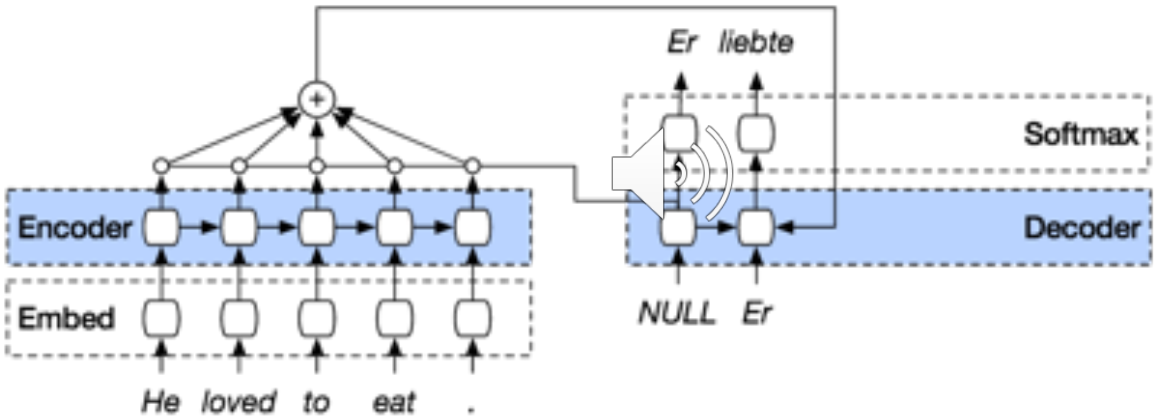
Julien Velcin

DIU EIL, 2019-2020

Traduire une langue en une autre



Réseaux de neurones profonds (*deep learning*)



source: https://smerity.com/articles/2016/google_nmt_arch.html

L’apprentissage automatique … en pratique

Julien Velcin

DIU EIL, 2019-2020

Quelques précautions à prendre

- Bien définir la tâche qu'on souhaite résoudre
- Identifier les données qui vont permettre à la machine d'apprendre (attention aux biais !)
- Préparer les données à l'apprentissage
- Apprendre à bien généraliser : méthodologie de l'apprentissage automatique et sur-apprentissage
- Souvent plusieurs critères : précision des résultats, interprétabilité, consommation et empreinte écologique...

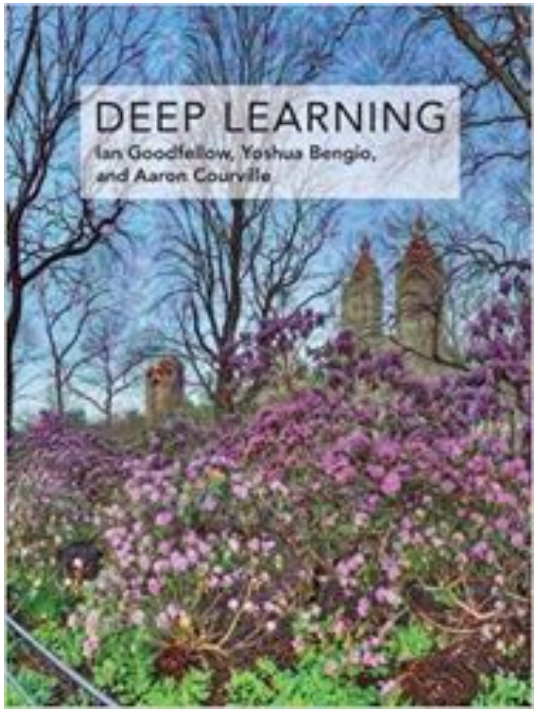
Quelques références

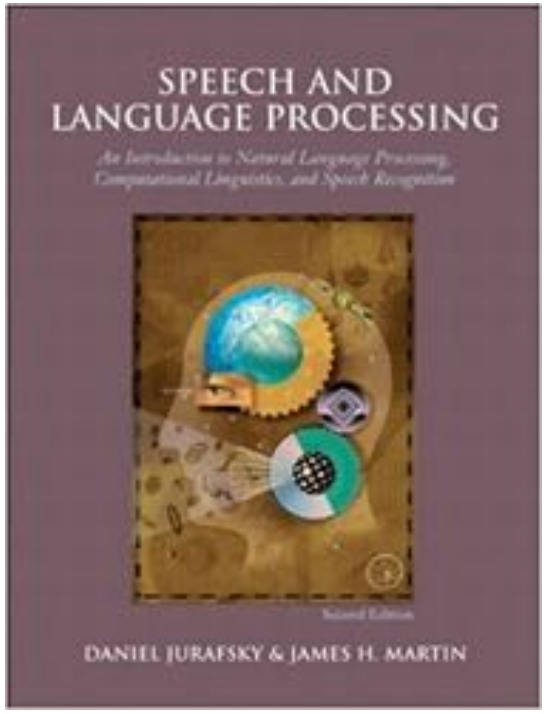
Julien Velcin

DIU EIL, 2019-2020

Ouvrages

3e édition
Antoine Cornuéjols
Laurent Miclet • Vincent Barra
Apprentissage artificiel
Deep learning, concepts et algorithmes
EYROLLES


DEEP LEARNING
Ian Goodfellow, Yoshua Bengio, and Aaron Courville


SPEECH AND LANGUAGE PROCESSING
DANIEL JURAFSKY & JAMES H. MARTIN