Can AI accurately predict forage energy and protein values using chemical and textual data ?

Aristide Lauront, Matéo Petitet, Raphaël Genin et Raphaël Rubrice

AgroParisTech

université PARIS-SACLAY

A project of the Association Française de Zootechnie (AFZ)
Supervised by Valérie Heuzé and Gilles Tran

Challenge : predicting forage value

Ration composition :

- impacts yield & economic revenue
- impacts animal health and welfare
- shall maximize quality/price ratio

French National Institute of Agricultural and Environmental Research (INRAE) helped develop nutrition system based on energy and protein intakes.

Why ?

- high variability
- right calculation is convoluted
- forage description might be complex

For what purpose ?

- creation of ration on the fly
- precise monitoring of nutrient and energy supply

Focus on **predictors** - Labels (1)

5 levels of semi-structured textual data

ID	Label 0	Label 1	Label 2	Label 3	Label 4
FV0230	FOURRAGES VERTS	PRAIRIES PERMANENTES, MONTAGNE (ALPES NORD)	À base de graminées	1er cycle	Épiaison du dactyle
FE0490	ENSILAGES	PRAIRIES PERMANENTES, PLAINE (NORMANDIE)	Préfané coupe fine	1er cycle	15-25 mai, début épiaison, ST = 470°C
FF0910	FOINS	GRAMINÉES FOURRAGÈRES, RAY-GRASS D'ITALIE, ALTERNATIF	Ventilé	1er cycle	Épi à 10 cm du sol

From level 0 to level 2 labels are consistent, above level two they are a lot more random

Type of data available (2)

Data from INRAE describing hundreds of **forages** characteristics. Each row made of :

ID	DM	Ashes	MAT	CF	NDF	ADF	EE	UFL	UFV	PDIA	PDI	BPR
FV0230	21,0	80	118	307	581	331	24	0,81	0,74	28	79	-10
FE0490	19,7	108	182	318	460	330	38	0,82	0,75	27	74	58
FF0910	85	130	219	234	532	271	38	0,96	0,93	51	104	65

Chemical values

6 chemical values taken in laboratory

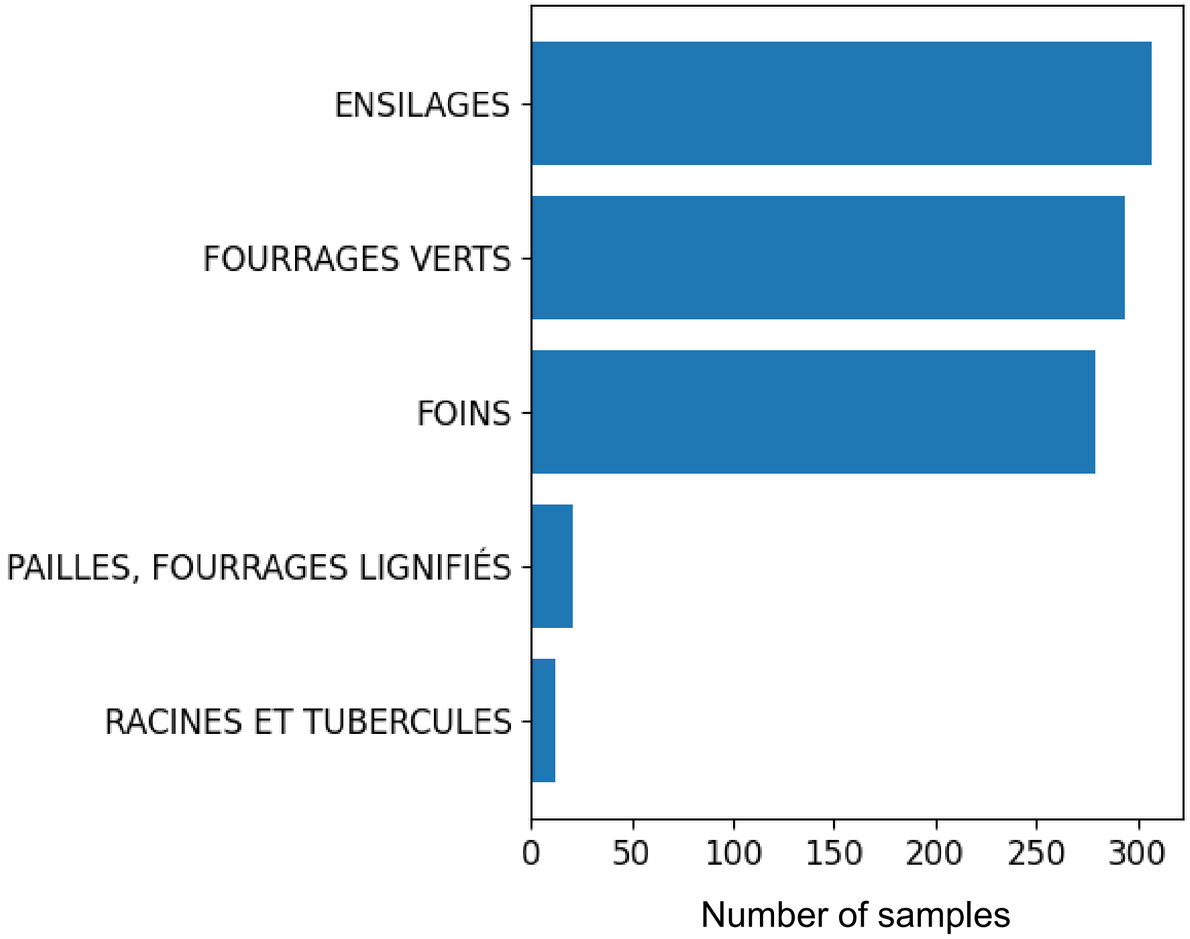
Nutritional values

Focus on predicted values

Calculated/defined by INRAE nutrition system

Energy	Protein
UFL : energy available for milking purpose (1 UFL = 1700 kcal)	**PDI** : Digestible proteins in the gut, nitrogen nutrient values (metabolizable proteins) in ruminants
UFV : energy available for cattle fattening (1 UFL = 1820 kcal)	**PDIA** : PDI of food origin, not degraded in the rumen
	BPR : protein balance of the rumen

Forage types

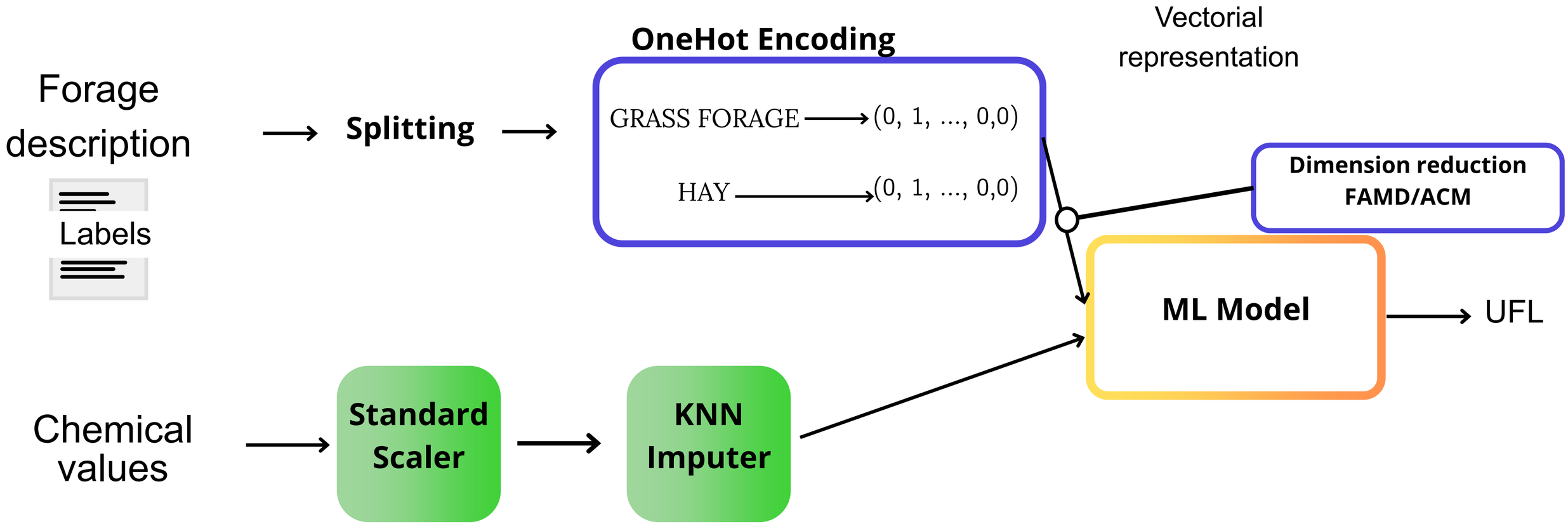


Tested approaches

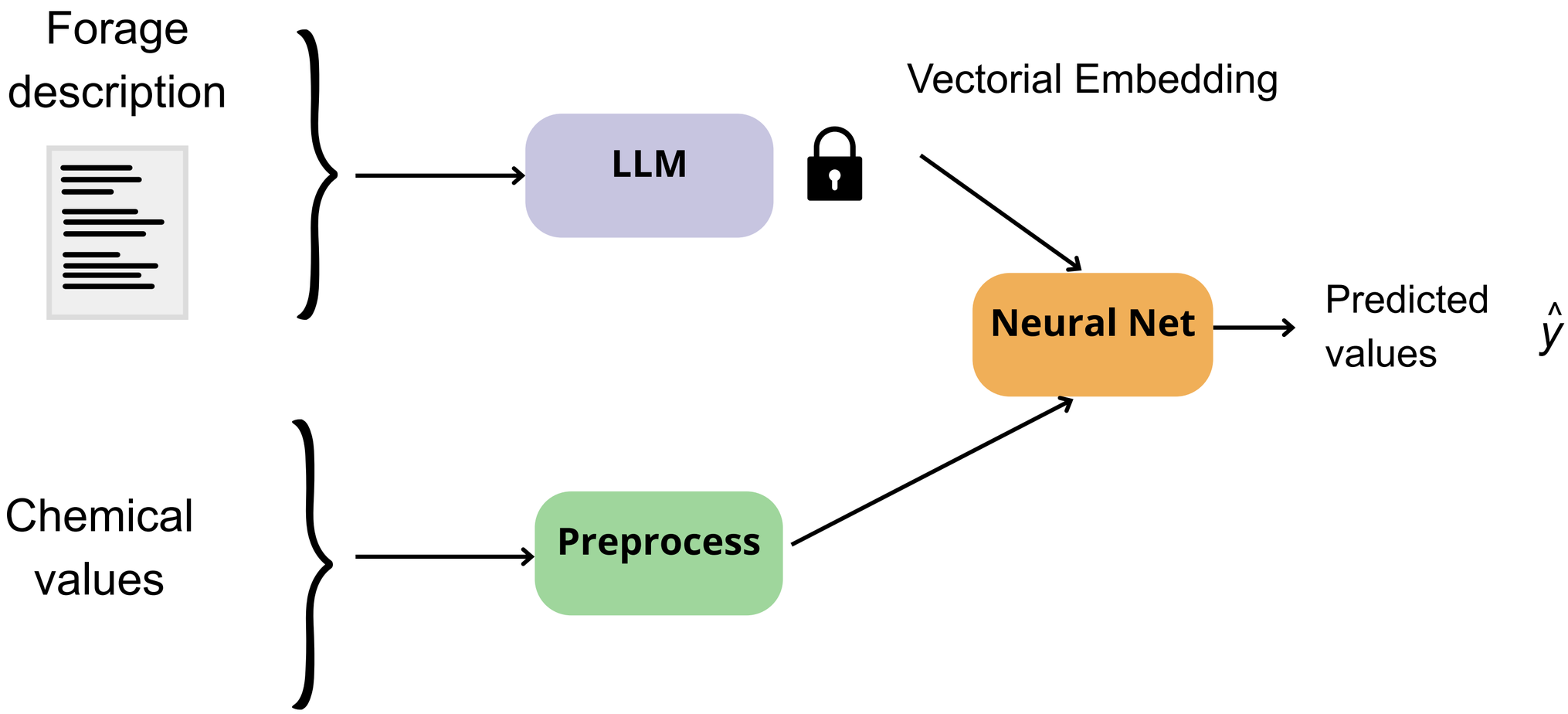
- First approach - Machine learning

- Second approach - LLM based

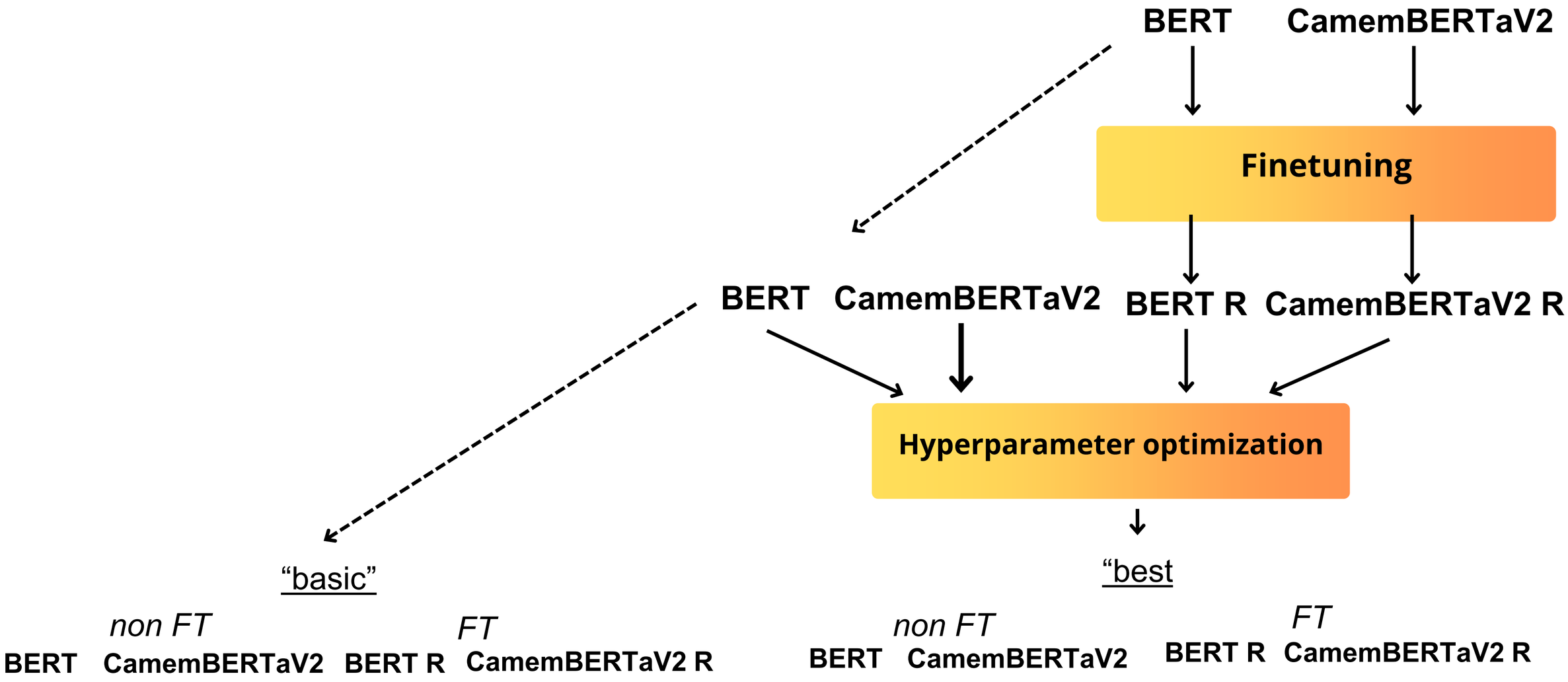
First approach - Machine learning



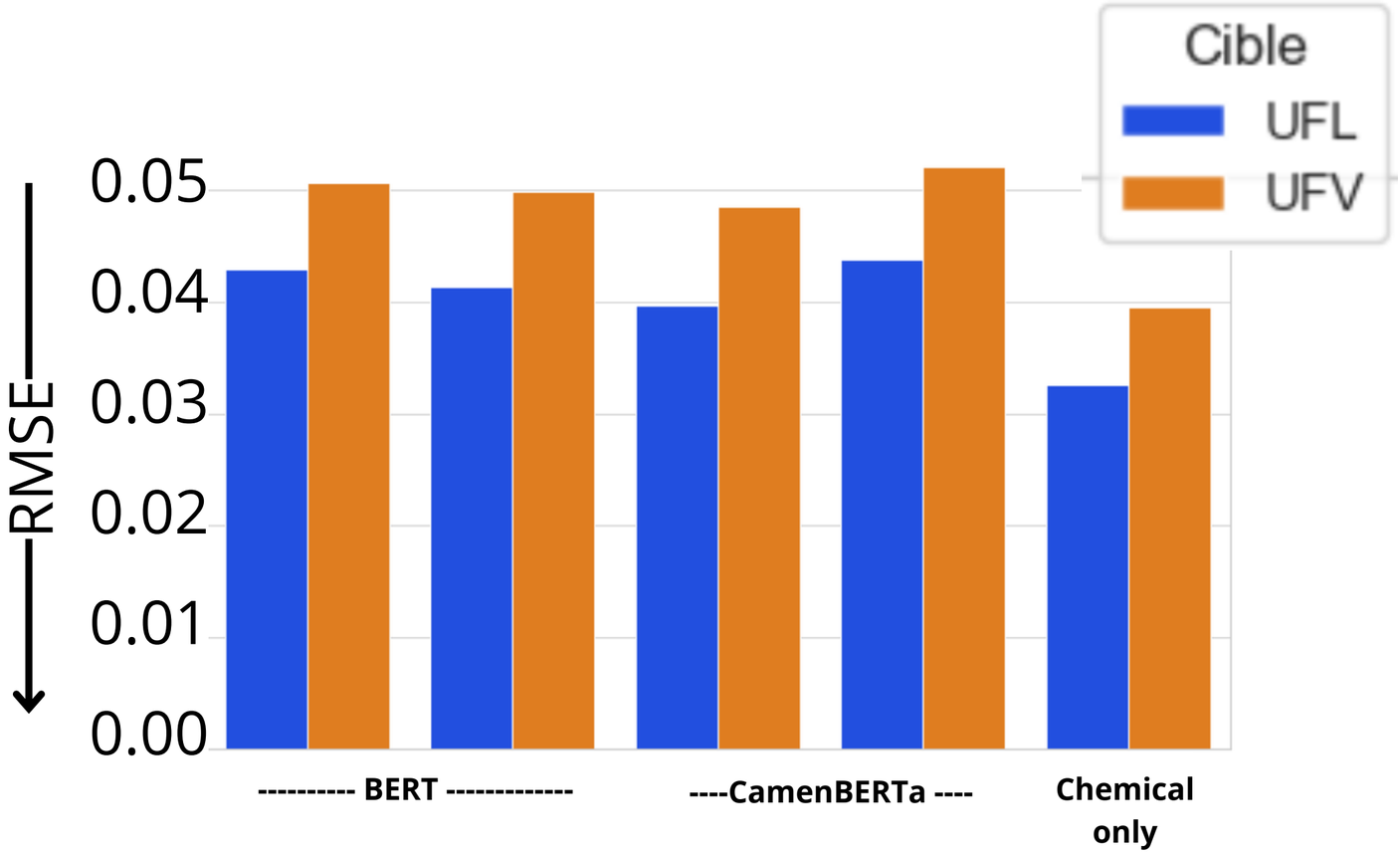
Second approach - LLM based



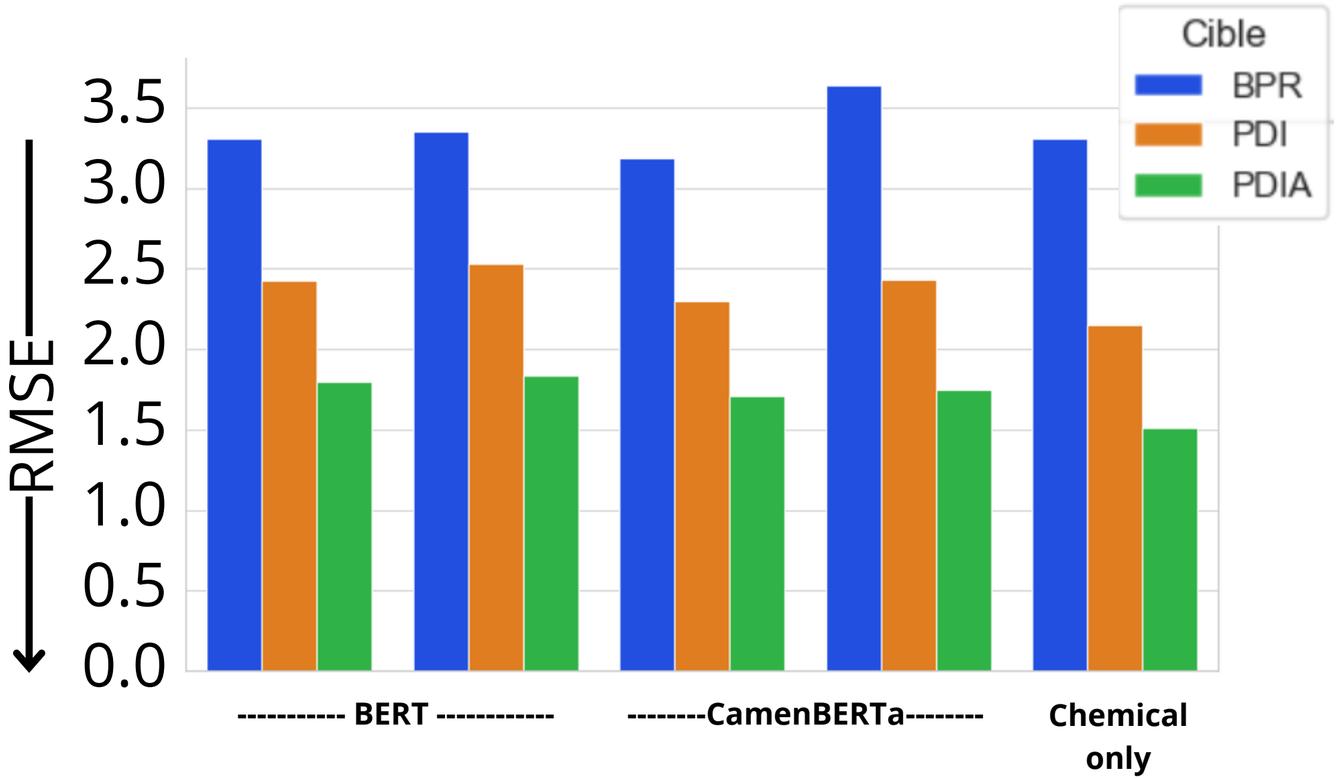
LLM - Model trained



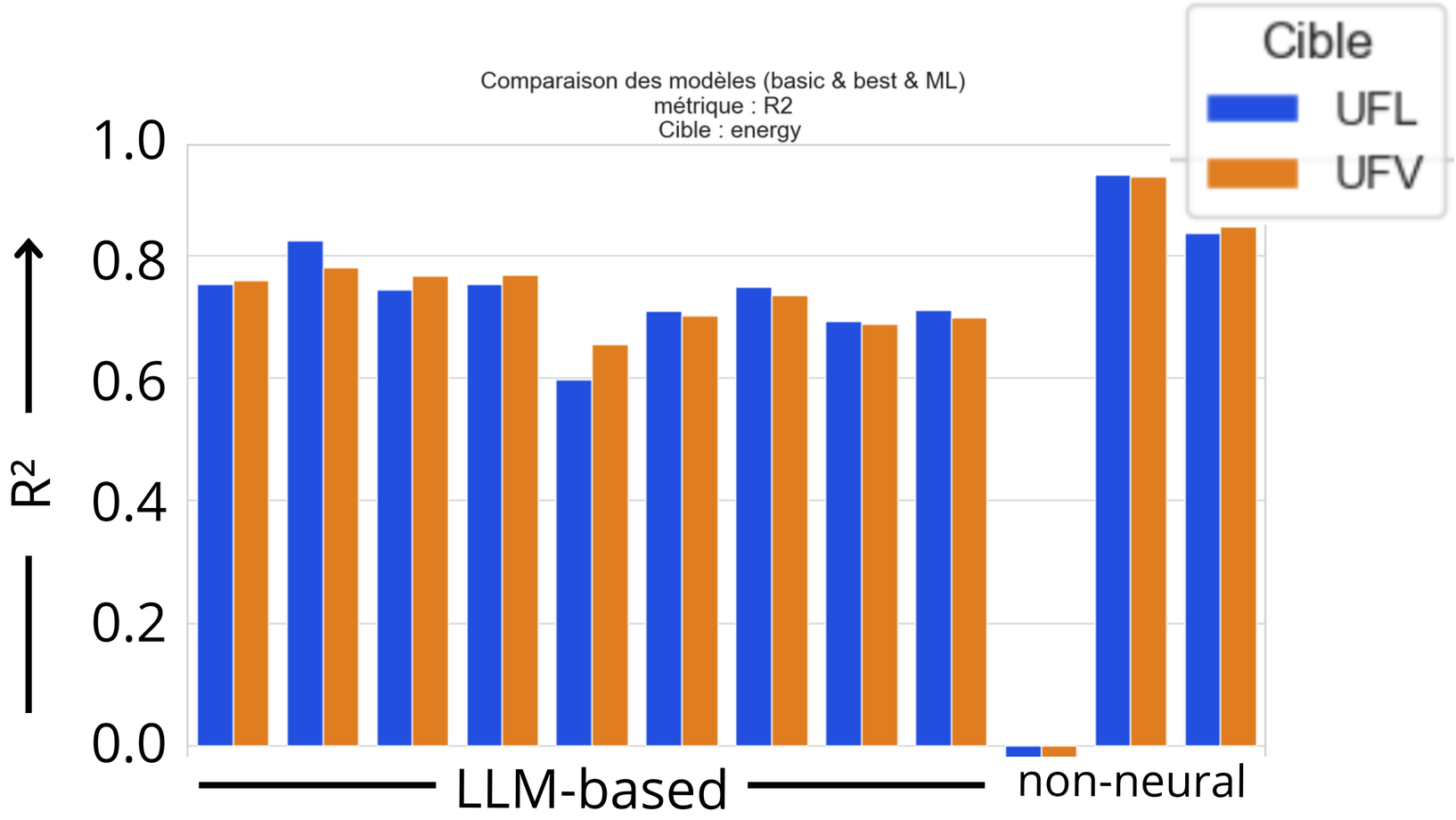
Results - Energy values



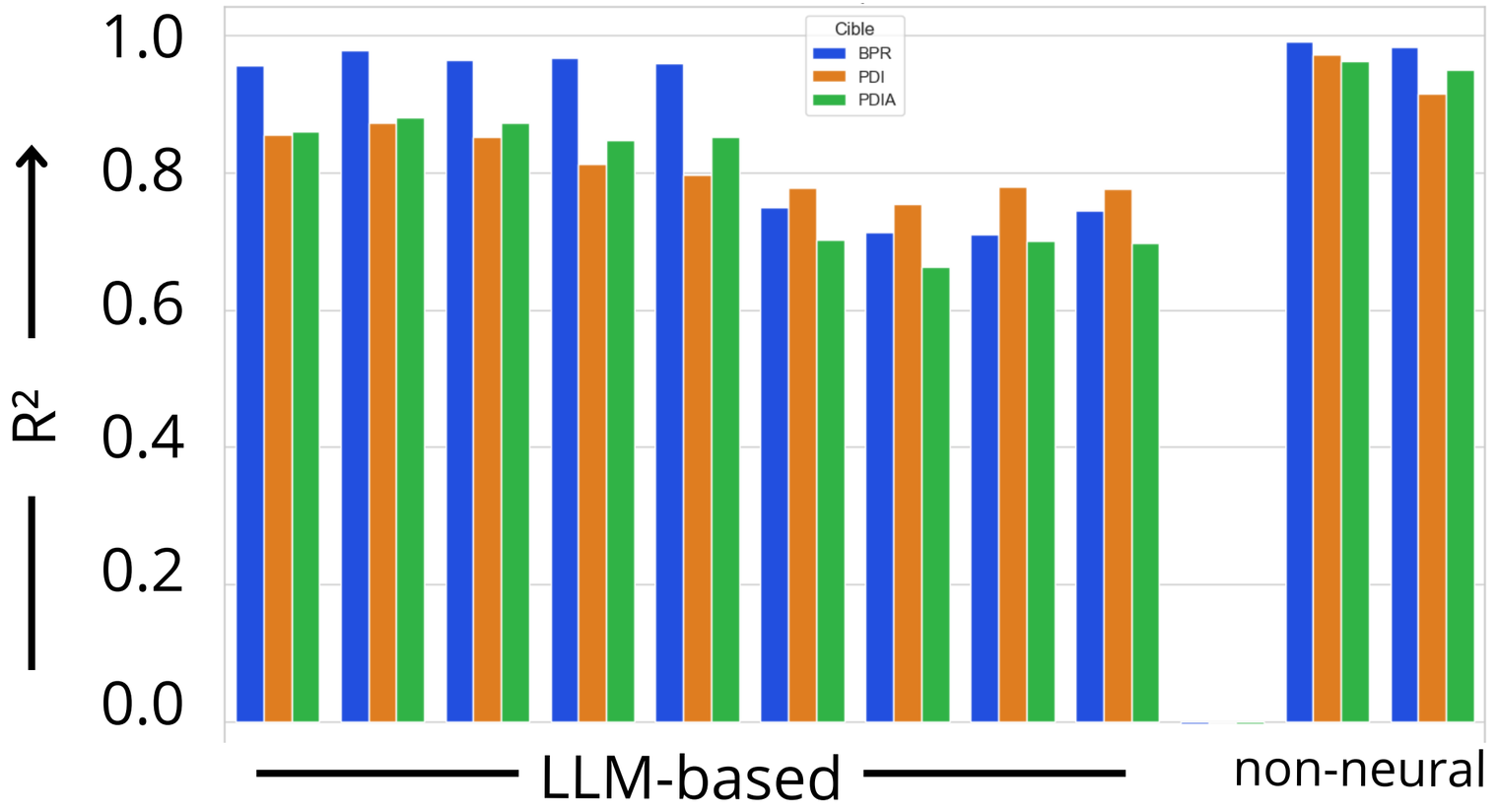
Results - Protein values



Results - Energy values



Results - Protein values



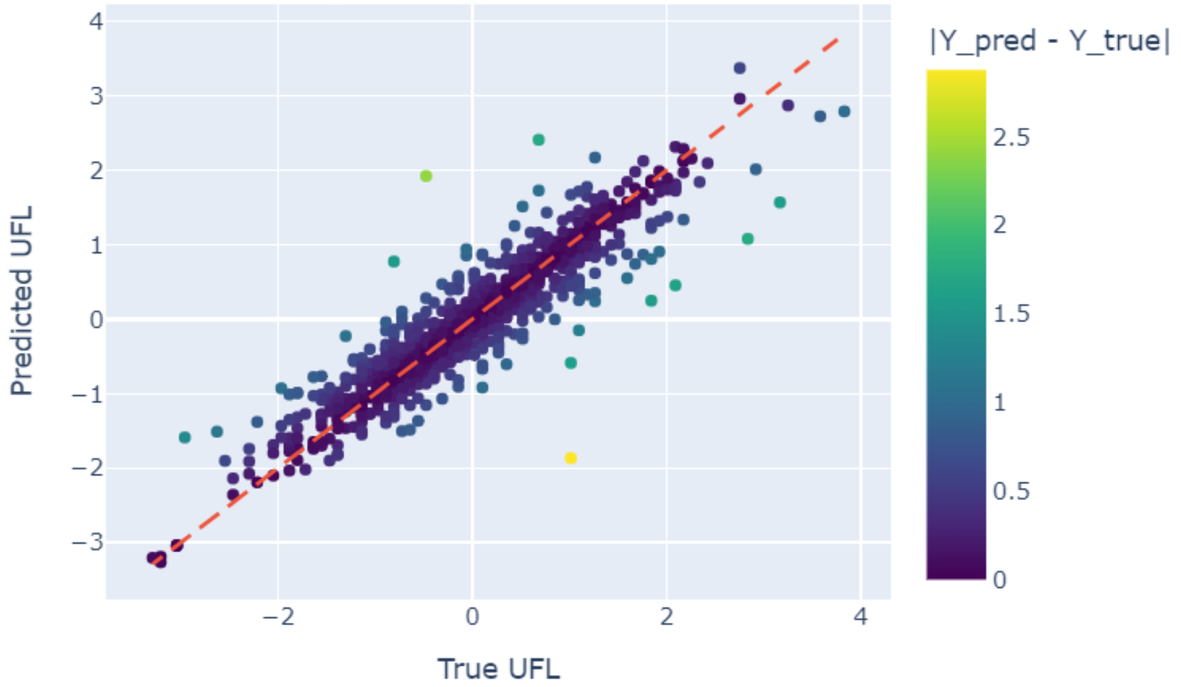
Key Takeaways

- Promising results for predicting all targets
- Best model : non-neural approach (XGBoost)
- Deep learning models with contrasting results:
  - comparable or slightly better results than non-neural
  - better generalization capacity
  - higher inference and training costs
- Prospects: field experimentation, use with routine systems, generalization to other fields

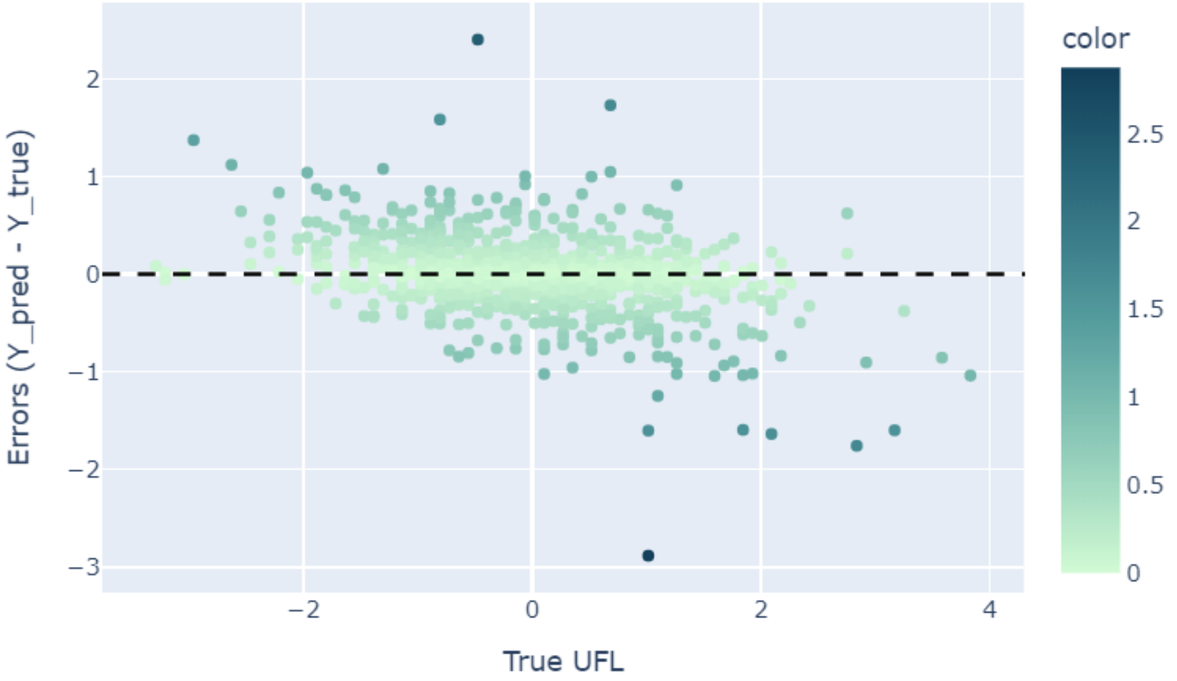
ADDITIONAL MATERIAL

Best model – non neural UFL

Meilleur Modèle pour prédire UFL: XGBoostRegressor
Dataset:MCA sur tous les Libellés + VC
MAE score: 0.2591, R² score: 0.8521



Plot des Résidus.
Ecart-type = 0.2842, Erreur Absolue Moyenne = 0.2591

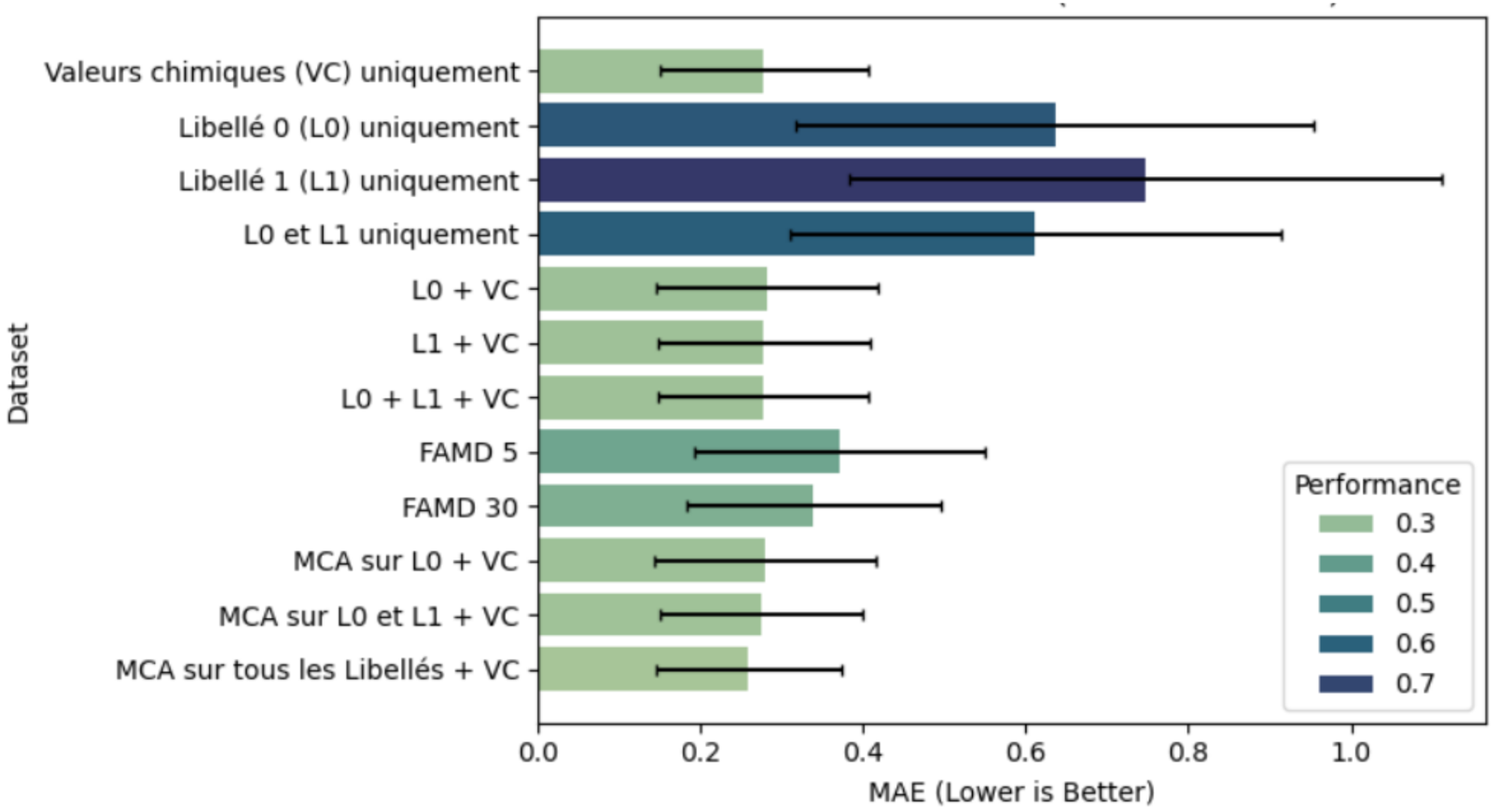


a

Non neural

Pre-process ?

UFL



b