Value of patient history for predictive models: a case study using French PMSI database

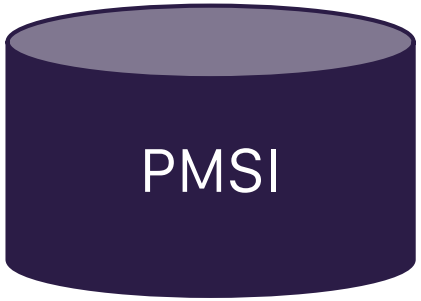
Hugo De Oliveira, Martin Prodel, Vincent Augusto

EURO 2018 – Valencia
11/07/2018

Contact : HDEOLIVEIRA@hevaeweb.com

PMSI database

The French National Hospital database



Medico-administrative database
→ *No electronic medical record*
→ *No biomedical test results*

Anonymization of Data

Hospital stay

- Created for hospital funding (activity based)
- Data available since 2006
- A very huge and complex structure
  (> 25 million hospital stays per year, many tables…)

Problematic

HEVA

Stroke in France

- 1st cause of handicap
- 2nd cause of dementias
- 2nd cause of mortality
- 130 000 cases per year
- Important aftermaths in 40% of cases

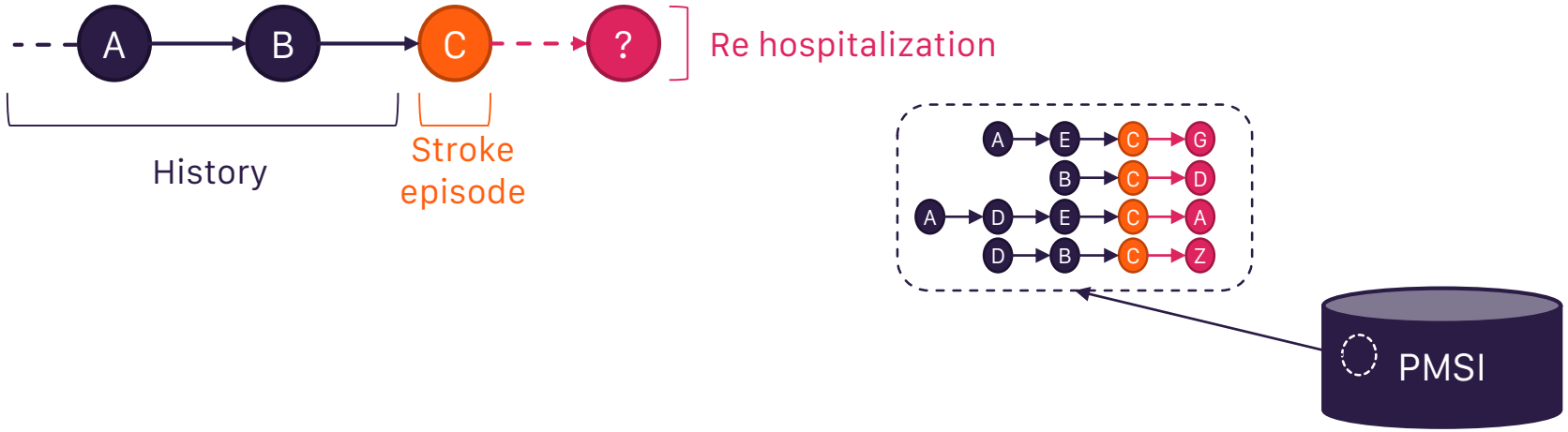
Study case

Prediction of patient's re-hospitalization stroke episode

INSERM, *Accident vasculaire cerebral*, 2013, https://www.inserm.fr/information-en-sante/dossiers-information/accident-vasculaire-cerebral-avc

Problematic

Objectives

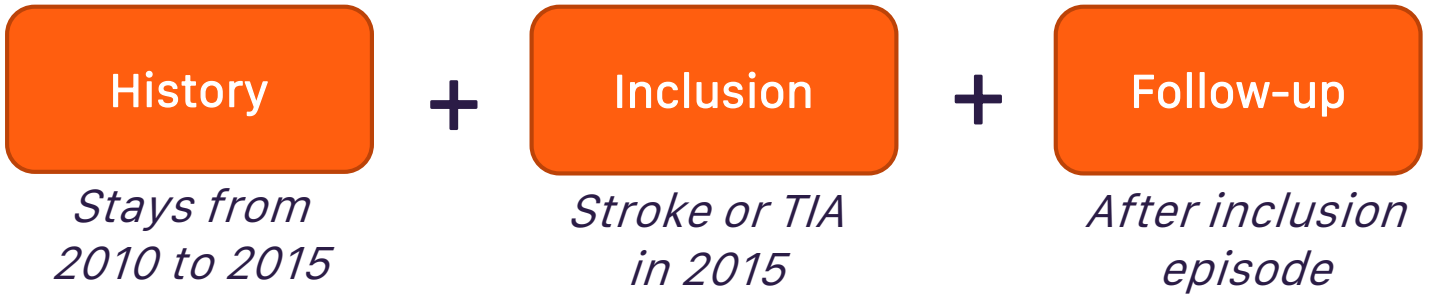


- A data processing method to take into account patient history with classical Machine Learning algorithms
- Test the impact of patient history in predictions

Data transformation

HEVA

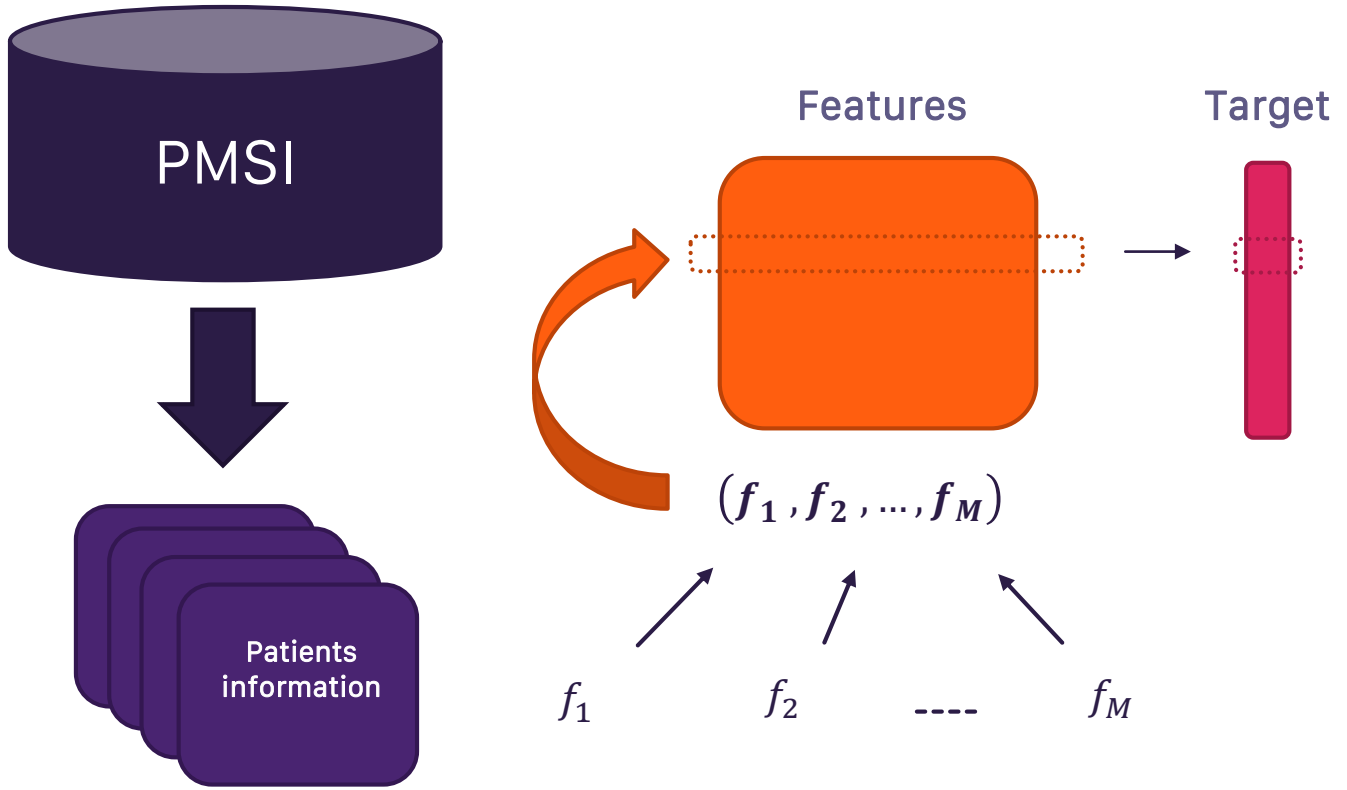
Study cohort



- **Inclusion :** patients with at least one stay with **Principal Diagnosis related to stroke or transient ischemic attack** (according to ICD-10 code) in **2015**
- **History :** analysis of all stays **from 2010** to the **first episode of stroke or TIA**
- **Follow-up :** analysis of **first re-hospitalization** of these patients (all diagnosis and cardiovascular diseases)

Data transformation

PMSI transformation to apply classical Machine Learning algorithms



Data transformation

Features creation

HEVA

Patient i

$(f_1^i, f_2^i, \ldots$ ... ... $\ldots, f_{M-1}^i, f_M^i)$

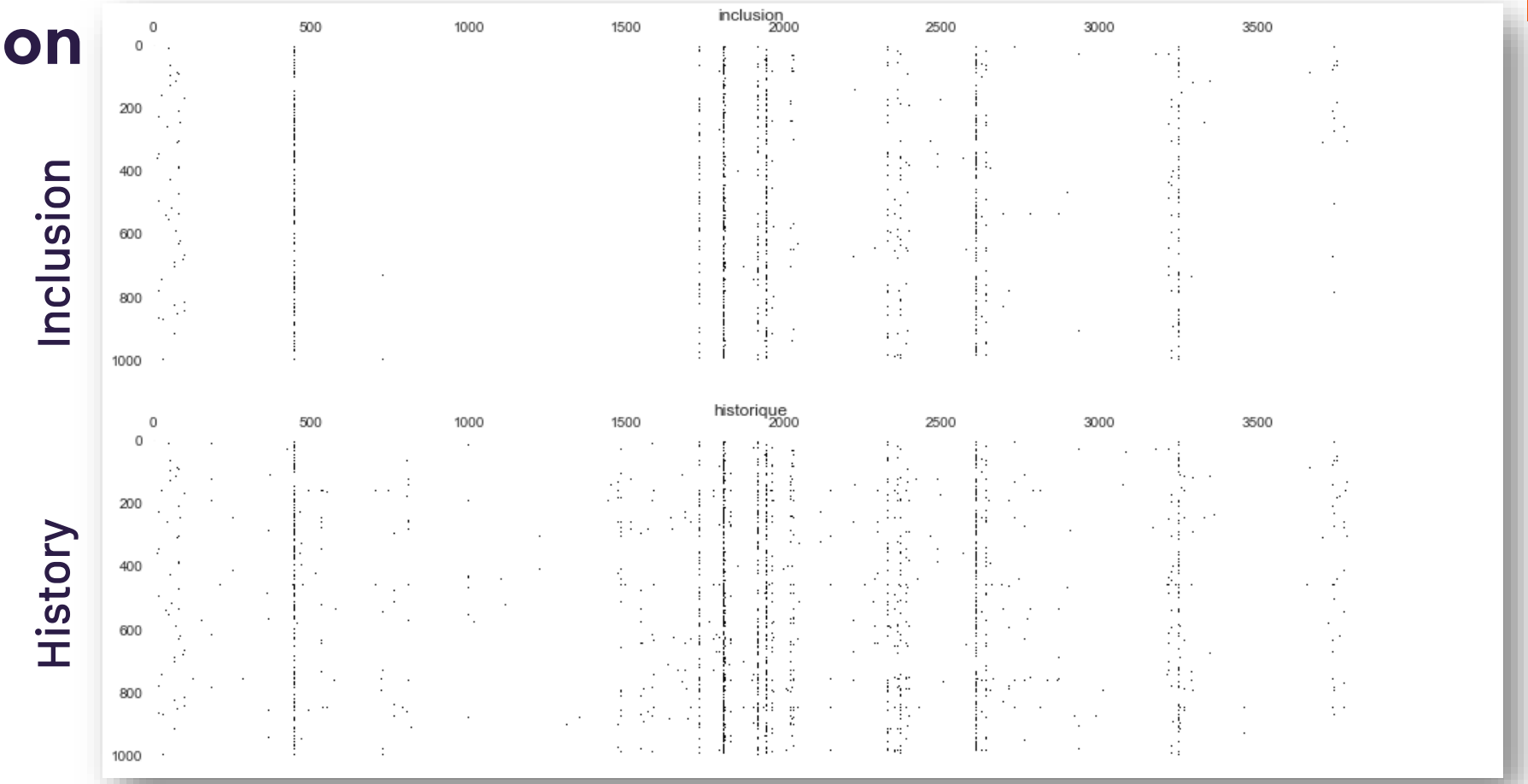
General Information	Principal diagnosis	Medical procedure	Related diagnosis	Medical unit information
Age, gender, city, duration of stay (inclusion)	*First 3 characters of the code*	*3d level of the coding hierarchy*	*First 3 characters of the code*	*Medical unit*

→ **89 011** patients
3 787 features

! Sparsity : more than 99% !

Data transformation

Sparsity

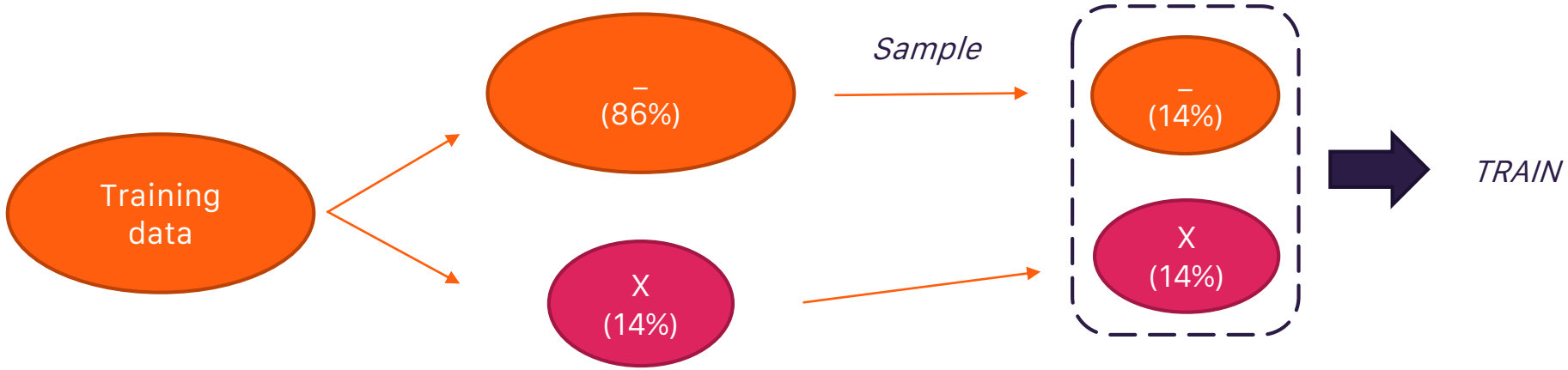


Representation of 1000 first rows of datasets (inclusion and history)

Data Balancing

Unbalance data problem

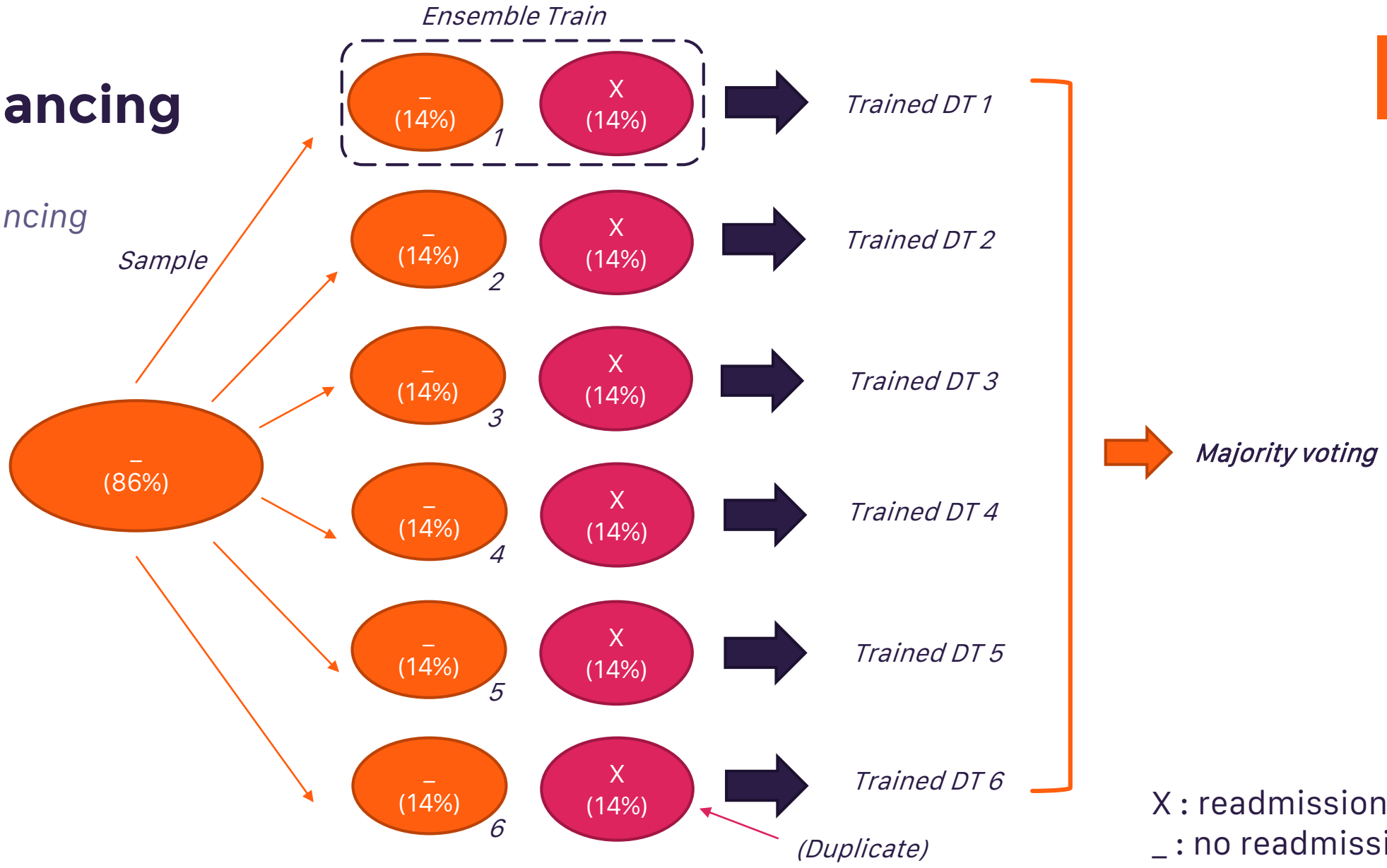
Data balancing



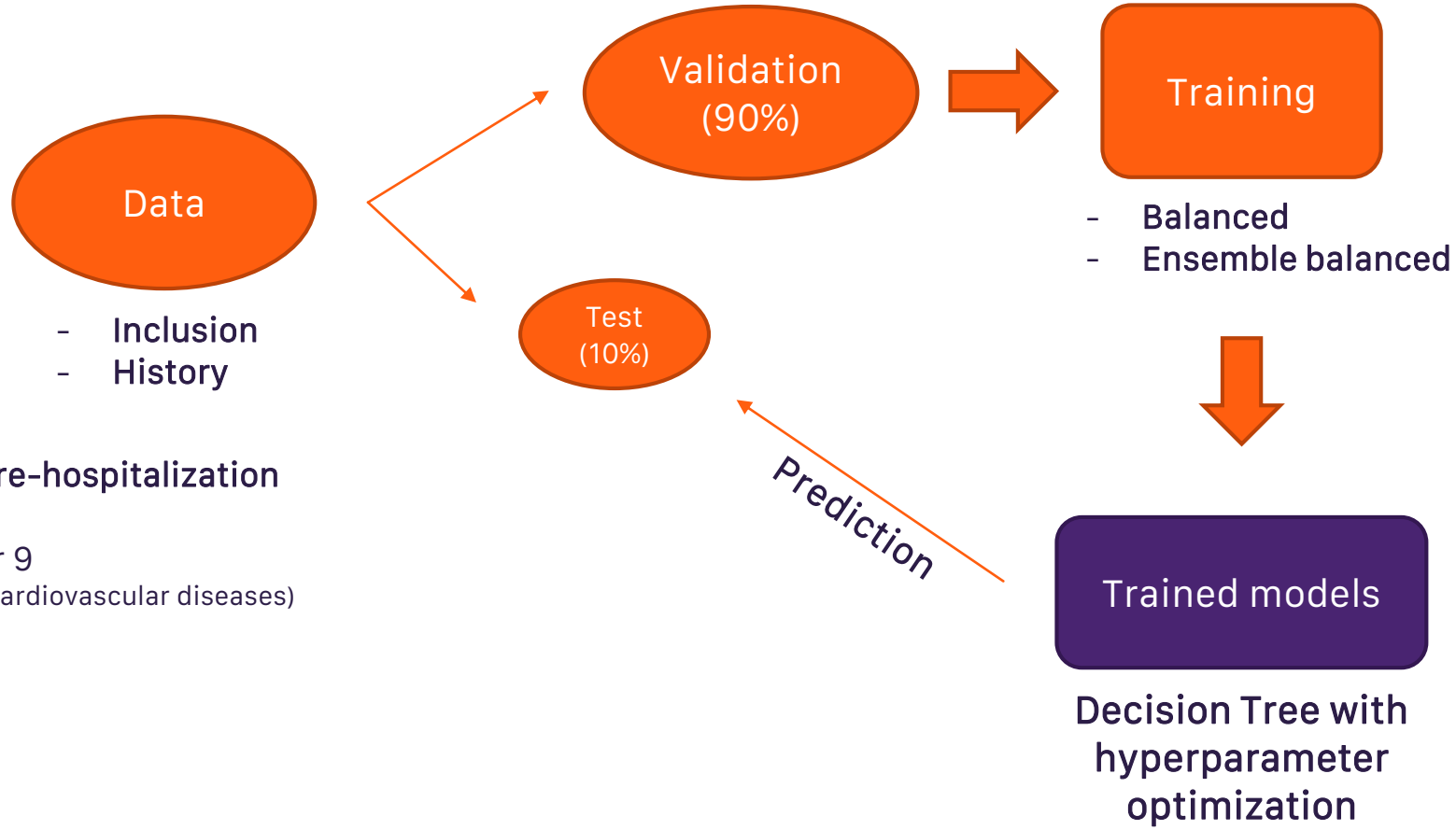
Ex : re-hospitalization within 30 days

Data Balancing

Ensemble balancing



Experiment plan



- Inclusion
- History

Prediction re-hospitalization
- General
- Chapter 9
  (ICD 10 : cardiovascular diseases)

Within
- 30 days
- 60 days
- 90 days

Scores
- Precision
- Recall
- F1-score

Results

Performances of prediction

		General admission						Chapter 9 diagnosis					
		Balanced			Ensemble balanced			Balanced			Ensemble balanced		
	Data	Precision	Recall	F1-score	Precision	Recall	F1-score	Precision	Recall	F1-score	Precision	Recall	F1-score
30 days	Inclusion	0,140	0,504	0,219	0,146	0,575	0,233	0,054	0,484	0,097	0,054	0,603	0,099
	Historique	0,137	0,665	0,227	0,152	0,599	0,242	0,056	0,455	0,100	0,058	0,564	0,105
60 days	Inclusion	0,221	0,499	0,306	0,222	0,565	0,319	0,086	0,533	0,148	0,086	0,639	0,152
	Historique	0,238	0,560	0,334	0,241	0,573	0,339	0,082	0,615	0,145	0,087	0,633	0,153
90 days	Inclusion	0,283	0,501	0,362	0,281	0,560	0,374	0,100	0,566	0,170	0,100	0,620	0,172
	Historique	0,289	0,546	0,378	0,299	0,586	0,396	0,105	0,667	0,182	0,105	0,672	0,182

- **History** improve performances in general
- **Ensemble balanced** improve performances too
- Prediction of relapse within a **longer period** after stroke improve **precision** but not really the **recall**

Conclusion & Perspectives

HEVA

Contributions

- A data processing method to take into account patient history with classical Machine Learning algorithms
- A proposed method for unbalanced data processing
- Illustration in a giver study case

Conclusions

- Adding history globally improve performances
- Results are not sufficient enough for practical use yet

Perspectives

- Development of new predictive models taking into account history and patient pathway

Thank you for your attention

Hugo De Oliveira
HDEOLIVEIRA@hevaweb.com

Annexes

Hugo De Oliveira
HDEOLIVEIRA@hevaweb.com

Results

Feature Importance

HEVA

TOP 5

Général

30 jours		180 jours	
inclusion	historique	inclusion	historique
duree_séjour	Acte rad. App. Resp.	duree_séjour	Acte rad. App. Resp.
das_C78	Radiologie	age_annee	Radiologie
code_res_31	duree_sejour	das_N18	das_I10
dp_I63	code_res_31	dp_G45	duree_sejour
age_annee	dp_G45	das_I48	code_res_31

Chapitre 9

Inclusion	historique	inclusion	Historique
das_I65	das_I65	das_I48	das_I48
dp_G45	dp_G45	das_I65	das_I65
age	dp_I50	dp_G45	dp_G45
das_I25	type_UM_02A	age_annee	Acte exp. Ap. circulatoire
duree_séjour	type_UM_02B	/	Acte rad. App. Resp.

General information

Inclusion

Occlusion et sténose (sans AVC)

Soins intensifs (cardio et autre)