

Résumé de l'exploration pour le Livrable 2

Variable à Inférer	Nombre de classe	Colonnes utilisées	Difficultés rencontrées	Approches testées	Approche retenues
Type de dispositif médical	1093	DESCRIPTION INCIDENT, FABRICANT, LIBELLE COMMERCIALE	Classes déséquilibrées, classes très rares, données avec des faux doublons rendant la séparation train/test difficile	Random Forest, Naïve Bayes, Linear SVM, Xgboost, Approche par similarité de vecteur embedding	Linear SVM
Type de dysfonctionnement	627	DESCRIPTION INCIDENT, ETAT PATIENT, ACTION_PATIENT et FABRICANT	Problème multi label, classes déséquilibrées, doublon non cohérent, part importante de NaN (25%)	SVM, XGBoost, NBSVM, LSTM, bi-LSTM, Attention, RNN, CNN, text-CNN, Xtreme Classification, RCNN, Zero-shot Learning, Camembert Vanilla	bi-LSTM
Conséquence du Dysfonctionnement	62	DESCRIPTION INCIDENT, ETAT PATIENT, ACTION_PATIENT et FABRICANT	Problème multi label, classes déséquilibrées, doublon non cohérent, part importante de NaN (60%)	SVM, XGBoost, NBSVM, LSTM, bi-LSTM, Attention, RNN, CNN, texte-CNN	bi-LSTM
Type d'effet	273	DESCRIPTION INCIDENT, ETAT PATIENT, ACTION_PATIENT et FABRICANT	Problème multi label, classes déséquilibrées, doublon non cohérent, part importante de NaN (30%)	SVM, XGBoost, NBSVM, LSTM, bi-LSTM, Attention, RNN, CNN, texte-CNN	bi-LSTM
Gravité (5 classe)	5	DESCRIPTION INCIDENT, ETAT PATIENT, CLASSIFICATION et FABRICANT	Classes extrêmement déséquilibrées	SVM, Random Forest, Naive Bayes, Ordinal SVM, LSTM	Ordinal SVM
Gravité (2 classe)	2	DESCRIPTION INCIDENT, ETAT PATIENT, CLASSIFICATION et FABRICANT	Classes extrêmement déséquilibrées	SVM, Random Forest, Naïve Bayes, Camembert	SVM