**Projet Deep Learning – DSTI**

**Participants :**

Alex Stephane Yeinko Tago

Hassane Kadri Dodo

Vincent Weber

**Question :**

Peut-on prévoir à partir de méthodes de DEEP LEARNING, NLP les prochaines activités médicales de patients dans le cadre d’une prise en charge complète de cancérologie ?

**Jeu de données disponible :**

Activité de patients enregistrée pendant 3 ans, (environ 500 patients) sur les années de 2019, 2020 et 2021, rassemblée dans un tableau « log » qui comporte les colonnes suivantes :

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| NIP | Date début | Date  fin | Code Acte | Service | Activité | Phase | Code Médecin |
|  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |

**Approche envisagée :**

1. **Feature Engineering** : Création d’une « phrase » de parcours patient

A partir des enregistrement de « log », créer une « phrase » correspondant au parcours d’un patient, intégrant l’ensemble des données du patient :

Par exemple : chaque ligne du jeu de donnée sera transformée en un code particulier intégrant les caractéristiques de l’acte réalisé.

L’aspect temporel sera implémenté dans le sens de lecture de la « phrase », le premier code correspond au premier acte du parcours de soins.

Les périodes sans activité seront matérialisées par un code spécifique matérialisant une non activité pendant une période donnée.

L’ordre des actes réalisés dans une même journée est un paramètre à considérer.

2 – **Création des training et test sets :**

De façon analogue à la réalisation des prévisions de Time Séries, il faut créer une base d’entrainement composée d’une longueur définie du parcours en vecteur X, puis de demander au modèle de deviner le prochain acte de la série, puis avancer pour chaque nouveau timestep du parcours.

**3 – Applications de modèles ML linéaires :**

* RF/ SVM / NN

**4 – Application d’un modèle de Deep Learning type NLP**

* Choix du modèle
* Entrainement
* Prédiction + Test + Crossvalidation

**5- Comparaison et conclusion**

SELECT TOP(1000)

T\_Actes.NIP as NIP,

T\_Actes.[DD\_A] as Date\_acte,

DATEDIFF('d', T\_Actes.[DF\_A], T\_Actes.[DD\_A]) as Duree,

T\_UFX.Service as Service,

T\_UFX.Activite as Activite,

T\_UFX.Phase\_Parcours as Phase,

T\_Actes.INX as Code\_Medecin,

T\_Code\_Acte.[Cle\_Encode\_acte] as Code\_Acte

FROM [ICO\_Activite].[dbo].[Tmp\_Acte\_Encoding] as T\_Code\_Acte,

[ICO\_Activite].[dbo].[Listing\_UF\_V3] as T\_UFX,

[ICO\_Activite].[dbo].[Tmp\_A\_Actes\_Table\_Analyse] as T\_Actes

WHERE T\_Code\_Acte.ID\_A=T\_Actes.ID\_A AND

T\_Actes.UFX=T\_UFX.UFX\_Code AND

T\_Actes.NIP IN (

SELECT [NIP]

FROM [ICO\_Activite].[dbo].[Mlflow\_Exp\_NIP\_Cl\_P1\_CPP]

WHERE [Cl\_P]=1

)

ORDER BY NIP asc