**Analyse de sentiments de tweets sur Azure**

**I Collecte des données**

**Utilisation de la plateforme Microsoft Azure**

**Utilisation du Service Azure Data Factory**

**Creation d’un DataFlow avec les caractéristiques suivants :**

**Tweet ayant le hastag #2024**

**Tweets les plus récents**

**Tweets en langue Françaises**

**Re-tweet exclus de la requête GET**

**Nombre de tweet = 100**

**Apres Exécution de la premiere partie du Data Flow , un JSON est enregistré en ayant les données brutes des 100 tweets dans le container RAW du datalake**

**II Nettoyage des données**

**Utilisation d’ADF et du DataFlow , plusieurs transformations :**

**Suppression des mentions , des hastags , adresses url , des saut de lignes , tabulation , des emojis irrelevant , et normalisation pour tout mettre en minuscules , etc…**

**Apres ces transformation , stockage du JSON dans un autre container nommé : PRE-CLEAN**

**Pour la suite dans l’énoncé il est mentionné d’utilisé Azure Databricks , l’ayant déjà utilisé lors du dernière Capstone , je me suis tourné vers un autre solution similaire pour l’explorer :**

**Azure Machine Learning .**

**NB : Avec peu de données le choix le plus judicieux : Azure Function**

**Cadre Big Data : Azure Databricks**

**Dans Azure ML , creation d’un script avec plusieurs fonctions :**

**Lecture du JSON dans PRE-CLEAN , utilisation** **d’unidecode pour standardiser les charactères spéciaux .   
  
Enregistrement du resultat dans le container CLEAN au format JSON**

**III Analyse de Sentiment**

**Creation d’un service Azure : AzureAI pour l’analyse de sentiment .   
Avec ce service plus besoin de faire la tokenisation et filtration des stops-words**

**Enregistrement du resultat sous forme de dataframe , sauvegardé dans un autre container : SentimentalAnalysis au Format Parquet ( plus léger que csv ) et compatible avec PowerBI**

**IV CI/CD**

**N/A**

**Le seul script est dans le notebook d’azure ML**

**V Visualisation**

**Graphique en Anneaux**

**Voir piece jointe**