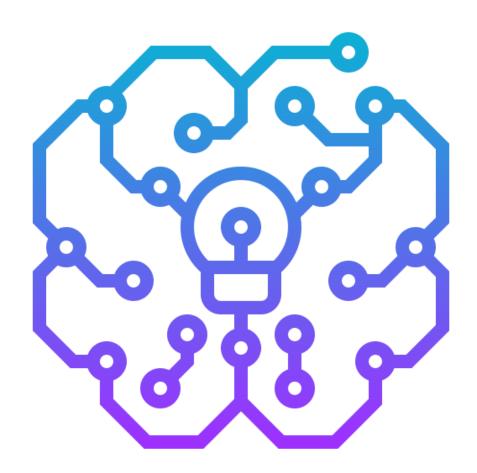




Machine Learning Project LSI 2023

Project 1: Arabic Automated short answers grading system for Moroccan history



Réalise par :

EL KRISSI Achraf AKDI Nafia

Encadre par:

Mr. EL AACHAK Lotfi

Table des matières

INTRODUCTION3
o CRISP
LA COMPRÉHENSION DU PROBLÈME MÉTIER4
LA COMPRÉHENSION DES DONNÉES5
LA CONSTRUCTION DU DATA HUB (PRÉPARATION DES DONNÉES)8
LA MODÉLISATION ET EVALUATION DES MODÈLES13
 MODÉLISATION
 EVALUATION
o RÉSUMÉ
 LA SAUVEGARDE DES MODÈLES
LE DÉPLOIEMENT DU MODÈLE24
o FastAPI
 Angular
o Graphql
MISE EN CONTENEUR AVEC DOCKER29
APPLICATION33

Introduction:

Le domaine de l'apprentissage automatique et de l'exploration de données offre des opportunités passionnantes pour extraire des connaissances précieuses à partir de vastes ensembles de données. Cependant, pour réussir ces projets, il est essentiel d'adopter une approche structurée et méthodique. C'est là que la méthodologie CRISP (Cross Industry Standard Process for Data Mining) entre en jeu.

La méthodologie CRISP fournit un cadre détaillé et itératif pour mener à bien des projets d'exploration de données et de modélisation prédictive. Elle est largement utilisée dans l'industrie et les domaines de recherche pour garantir une approche cohérente tout au long du processus. Dans ce rapport, nous allons explorer les différentes étapes de la méthodologie CRISP et les appliquer à notre projet de machine learning.

La première étape de la méthodologie CRISP consiste à comprendre le problème et les objectifs. Il est essentiel de définir clairement le problème à résoudre et d'identifier les objectifs commerciaux ou opérationnels sous-jacents. Cela nous permettra de cibler nos efforts de modélisation et de fournir des résultats pertinents pour la prise de décision.

La deuxième étape est la compréhension des données. Nous devons collecter les données pertinentes pour notre projet et les explorer en détail. Cela nous permettra de comprendre la qualité, le format et la structure des données, ainsi que d'identifier d'éventuels problèmes ou erreurs.

La troisième étape est la préparation des données. Cette étape implique le nettoyage, la transformation et la préparation des données pour l'analyse ultérieure. Nous devons traiter les valeurs manquantes, les données aberrantes et les variables catégorielles, et

normaliser les données si nécessaire. Une préparation minutieuse des données est essentielle pour garantir des résultats fiables et précis.

La quatrième étape est la modélisation des données. Nous appliquerons des techniques d'apprentissage automatique et de modélisation prédictive pour construire des modèles statistiques. Selon les objectifs de notre projet, nous pouvons utiliser des méthodes telles que la régression, la classification ou le clustering. Nous évaluerons également la performance de nos modèles pour sélectionner le meilleur modèle possible.

dans la cinquième étape, nous évaluerons les de notre modèle. Nous examinerons les métriques d'évaluation appropriées pour évaluer la performance de notre modèle. Si les résultats sont satisfaisants, nous pourrons déployer le modèle en production pour une utilisation continue.

Le déploiement est l'étape finale du processus. Elle consiste en une mise en production pour les utilisateurs finaux des modèles obtenus. Son objectif : mettre la connaissance obtenue par la modélisation, dans une forme adaptée, et l'intégrer au processus de prise de décision.

Le déploiement peut ainsi aller, selon les objectifs, de la simple génération d'un rapport décrivant les connaissances obtenues jusqu'à la mise en place d'une application, permettant l'utilisation du modèle obtenu, pour la prédiction de valeurs inconnues d'un élément d'intérêt.

La compréhension du problème métier :

Dans le cadre de ce projet, notre objectif est de développer un système automatisé de notation des réponses courtes en arabe pour l'histoire marocaine. L'objectif principal est de fournir une évaluation précise et adéquate aux étudiants en fonction de leurs réponses. Nous souhaitons évaluer leur compréhension des concepts

historiques, leur capacité à analyser les informations pertinentes et à formuler des arguments cohérents.

Grâce à ce système, nous visons à améliorer le processus de notation des réponses courtes en offrant une solution automatisée qui permettra aux enseignants de gagner du temps et de fournir des commentaires plus précis et constructifs aux étudiants. Nous cherchons à créer un environnement d'apprentissage plus efficace et interactif, où les étudiants pourront recevoir un retour rapide sur leurs performances et disposer d'un outil qui les aidera à comprendre leurs forces et leurs faiblesses dans la matière.

En comprenant pleinement le problème métier, nous serons en mesure de développer un système qui répondra aux besoins spécifiques des enseignants et des étudiants. Nous chercherons à fournir une évaluation objective et cohérente, tout en tenant compte des particularités linguistiques et culturelles propres à l'arabe et à l'histoire marocaine.

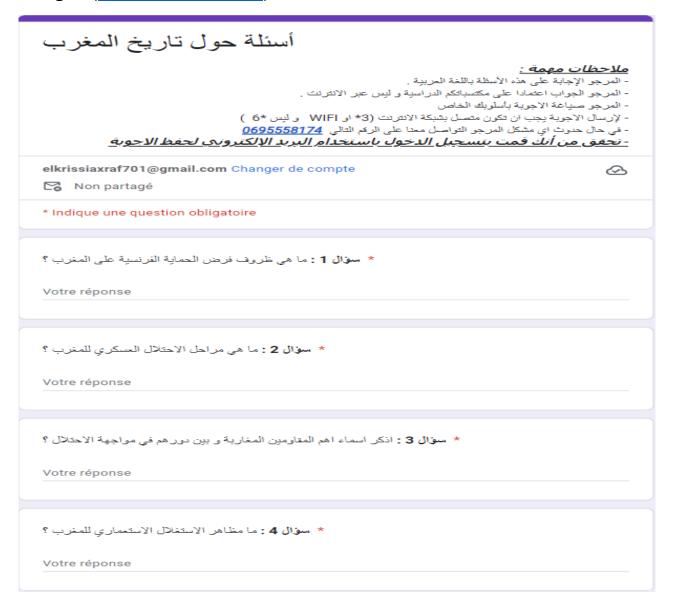
En résumé, notre objectif est de développer un système automatisé de notation des réponses courtes en arabe pour l'histoire marocaine qui fournira une évaluation précise et équitable aux étudiants, tout en offrant aux enseignants un outil efficace pour évaluer les performances des étudiants et leur fournir des commentaires constructifs.

La compréhension des données

Dans ce projet, nous avons mis en place un processus pour évaluer les élèves à travers 10 questions portant sur l'histoire du Maroc. Ces questions ont été soigneusement sélectionnées par un **ancien**

enseignant du lycée Hassan II à Tanger, afin de garantir leur pertinence et leur adéquation avec le programme d'études.

Pour collecter les réponses des élèves, nous avons créé un formulaire en ligne (aller au formulaire).



```
* سؤال 5 : ما هي انعكاسات الاستغلال الاستعماري على الاقتصاد و المجتمع المغربيين ؟
Votre réponse
* سؤال 6 : ما هي ظروف ظهور الحركة الوطنية ؟
Votre réponse
* سوال 7 : ما هي ظروف تقديم المغرب لوثيقة الاستقلال ؟
Votre réponse
* سؤال 8: ما علاقة ثورة الملك و الشعب باستقلال المغرب؟
Votre réponse

    * سؤال 9: ما هي الضغوط العسكرية التي استعملتها الدول الامبريالية للتغلغل الاوروبي في المغرب خلال القرن 19

ج ؟
Votre réponse
* سؤال 10: ما نوعية الاصلاحات التي باشرها المغرب خلال القرن 19 م و ما مآلها؟
Votre réponse
Envoyer
                                                                                 Effacer le formulaire
```

Ce formulaire permet aux élèves de répondre aux questions de manière structurée et organisée. Chaque réponse est automatiquement enregistrée dans un fichier Excel dédié, ce qui facilite la gestion et l'analyse des données.

Nous avons également essayé d'évaluer la qualité des réponses en se basant sur la réponse parfaite donnée par l'enseignant . Chaque réponse reçoit un score en fonction de sa précision, de sa pertinence et de la qualité de son argumentation. **Un score de 0** est attribué si la réponse est incorrecte, **un score de 1** est donné si la réponse est

correcte mais manque d'argumentation solide, et **un score de 2** est attribué si la réponse est parfaite, avec une argumentation claire et convaincante.

Α	В	C
number	value	score
	ظروف خارجية : التنافس الإمبريالي على المغرب نتج عنه توقيع اتفاقات انتهت بأنفراد فرنسا واسبانيا باحتلال المغرب الاتفاق الفرنسي الإيطالي 1912حول مغرب وليبيا الاتفاق الفرنسي الألماني 1911حول مغا	2
	لظروف التي أدت إلى فرض الحماية الفرئسية على المغرب تشمل عوامل خارجية مثل المنافسة الإمبريالية بين القوى الأوروبية ، والتي أدت إلى توقيع اتفاقيات منحت فرنسا وإسبانيا احتلار احتلال المغرب. وتلأ	2
	برجع فرض الحماية الفرنسية على المغرب إلى عدة أسباب، منها الظروف الخارجية كالتنافس الإمبريالي بين فرنسا وإسبانيا وتوقيع اتفاقات احتلال المغرب، والاتفاق الفرنسي الإيطالي عام 1902 حول المغرب	
	وليبيا والاتفاق الفرنسي الألماني في 1911 حول المغرب والكونغو، وكذلك الظروف الداخلية مثل توالي السلاطين وتزايد عدد النورات والتمردات، وتزايد حجم القروض الأجنبية وإنقال كاهل الفلاحين	
	بالضرائب، وانتشار الفقر والأمراض والأوبئة والبطالة والمجاعة.	2
	ن فرض الحماية الفرنسية على المغرب يعود إلى عدة أسباب، منها الأسباب الخارجية التي تتمثل في التنافس الإمبريال بين فرنسا وإسبانيا وتوقيع اتفاقات لاحتلال المغرب، بما في ذلك الاتفاق الفرنسي الإيطالي [1]	
	فرض الحماية الفرنسية على المغرب بعود إلى عدة أسباب، بما في ذلك الضغوطات الخارجية مثل التنافس الإمبريالي بين فرنسا وإسبانيا والاتفاقات التي أدت إلى احتلال المغرب، وكذلك الظروف الداخلية مثل الوالم المعرب وكذلك الظروف الداخلية مثل التنافس الإمبريالي بين فرنسا وإسبانيا والاتفاقات التي أدت إلى احتلال المغرب، وكذلك الظروف الداخلية مثل الوالم المعرب والأمراض والأوبئة والبطالة والمجاعة.	
	رجع أسباب فرض الحماية الفرنسية على المغرب إلى مجموعة من العوامل، بما في ذلك التنافس الإمبريالي بين فرنسا وإسبانيا والاتفاقات التي أدت إلى احتلال المغرب، وكذلك التحديات الداخلية مثل توالي لسلاطين وتصاعد الثورات والتمردات وانتشار الفقر والأمراض والأوبئة والبطالة والمجاعة.	
	رجع فرض الحماية الفرنسية على المغرب إلى عدة عوامل، منها التنافس الإمبريالي بين فرنسا وإسبانيا وتوقيع الاتفاقات التي أدت إلى احتلال المغرب، والتحديات الداخلية مثل توالي السلاطين وتصاعد لثورات والتمردات وانتشار الفقر والأمراض والأوبئة والبطالة والمجاعة.	
	عدة أسباب دفعت فرنسا لفرض الحماية على المغرب، ومنها الظروف الخارجية كالتنافس الإمبريالي مع إسبانيا وتوقيع اتفاقات احتلال المغرب بما في ذلك الاتفاق الفرنسي الإيطالي لعام 1902 حول المغرب وليونية والبطالة والمجاعة. 191 حول المغرب والاكتفاق الفرنسي الألماني لعام 1911 حول المغرب والكونغو. ومن الأسباب الداخلية، تزايد عدد الثورات والتمردات وتوالي السلاطين وانتشار الفقر والأمراض والأويئة والبطالة والمجاعة.	
	نوقيع اتفاقات بالاحتلال الفرنسي والإسبائي للمغرب كان نتيجة للتنافس الإمبريالي على المغرب، وتم توقيع اتفاق الفرنسي الإيطالي في عام 1902 بشأن المغرب واليبيا، واتفقت فرنسا مع ألمانيا في عام 1911 على المغرب والكونغو. ومن الجوانب الداخلية، ضعفت السلطة المركزية بعد وفاة حاجب السلطان باحماد في عام 1900، وزادت عدد الثورات والتمردات بسرعة نتيجة تعاقب السلاطين، مثل تمرد الريسوني والجيلالي الزرعوني. وزاد حجم القروض الأجنبية، وتحمل الفلاحون الكثير من الضرائب والأعباء، مما أدى إلى انتشار الفقر والأمراض والأوبئة والبطالة والمجاعة.	
	الخلال الوينسي والرسائي تشغرب جاء الهدة تنتقص الإسرياني من الوينس من فرقيح على الوين وقد توقيع على الوين المناس المناس على عدر 1900 جل المناس وليناء كما أم الانتها على المنزس وليناء كما أم الانتها على المنزس الوينة ويمال اللاحية وحمل الانتجيز الكلام المناس الأخية وحمل الانتجيز الكلام المناس الم	
	لتنافس الإمبريالي بين فرنسا وإسبانياكان من بين الأسباب التي أدت إلى فرض الحماية الفرنسية على المغرب وتوقيع اتفاقات احتلال المغرب، بما في ذلك الاتفاق الفرنسي الإيطالي عام 1902 حول المغرب وليبيا والاتفاق الفرنسي الألماني في 1911 حول المغرب والكونغو. ومن الجوانب الماخلية، تزايدت الثورات والتمردات في ظل توالى السلاطين، مثل تمرد الريسوني والجيلالي الزرعوني، وتزايد حجم القروض لأجنبية والضرائب الثقيلة التي حملت الفلاحين الكثير منها، مما أدى إلى انتشار الفقر والأمراض والأوبئة والبطالة والمجاعة.	

La construction du Data Hub (préparation des données) :

```
# read csv file into pandas
df = pd.read_excel("quest10.xlsx")
df.head(10)
```

```
number
                                                         value score
0
        10.0
                              أقام المغرب مدارس جديدة لتعليم الفتيات
                                                                        2
        ...المغرب باشر إصلاحات اجتماعية لتحسين الوضع الاق 10.0
1
                                                                        1
2
        10.0
                    ...المغرب لم يقم بأي تغييرات اجتماعية يستحق الذكر
                                                                        0
3
        10.0
                   قام المغرب بإنشاء مستشفيات لتحسين الوضع الصحى
                                                                        2
        10.0
                        فتح المغرب بعض المدارس لتعليم العلوم الحديثة
                                                                        1
5
        10.0
                  ... المغرب لم يقم بأي إصلاحات اجتماعية خلال القرن
                                                                       0
6
        10.0
                      ... أنشأ المغرب بيوت الأيتام ودور الرعاية للفقراء
                                                                        2
                          نشر المغرب الوعى الصحى والتثقيف الصحى
7
        10.0
                                                                        1
                    ...المغرب لم يقم بأي تغييرات اجتماعية يستحق الذكر
8
        10.0
                                                                        0
9
        ... عمل المغرب على تحسين حماية حقوق العمال ونظام ا
                                                                       2
```

```
# infos sur notre dataset
df.info()
# les valeurs manquantes
df.isnull().sum()
# On remarque qu'il n'y a pas de valeurs nulles
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 97 entries, 0 to 96
Data columns (total 3 columns):
    Column Non-Null Count Dtype
 0
    number 97 non-null
                             int64
 1
    value
            97 non-null
                             object
     score
             97 non-null
                             int64
dtypes: int64(2), object(1)
memory usage: 2.4+ KB
number
          0
value
          0
score
          0
dtype: int64
```

```
# les valeurs dupliquées
df.duplicated().sum()
```

0

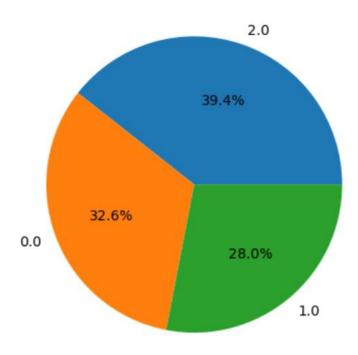
```
# Visualisation
score_counts = df['score'].value_counts()

# Créer le pie chart pour la répartition des scores
plt.pie(score_counts, labels=score_counts.index, autopct='%1.1f%%')

# Ajouter un titre
plt.title('Répartition des reponses')

# Afficher le graphique
plt.show()
```

Répartition des reponses



```
score_counts = df['score'].value_counts()

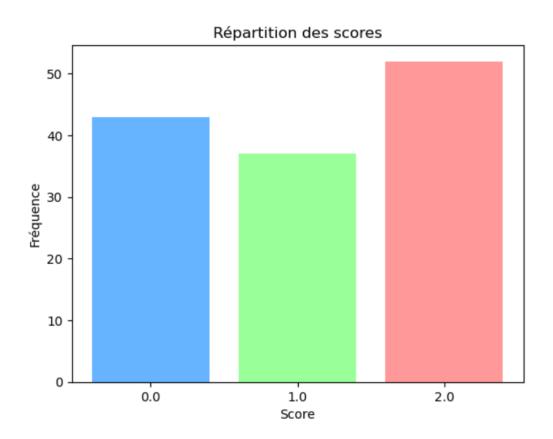
# Créer une liste de couleurs pour chaque score
colors = ['#ff9999', '#66b3ff', '#99ff99']

# Créer le bar chart avec des couleurs personnalisées
plt.bar(score_counts.index, score_counts.values, color=colors)

# Ajouter des étiquettes pour chaque valeur du score
plt.xticks(score_counts.index, score_counts.index)

# Ajouter un titre et des labels d'axe
plt.title('Répartition des scores')
plt.xlabel('Score')
plt.ylabel('Fréquence')

# Afficher Le graphique
plt.show()
```



On peut dire que notre data est balancé (données équilibrées)

```
stop_words = set(stopwords.words('arabic'))
def preprocess_text(text):
    text = text.lower()
    tokens = word_tokenize(text)
    #tokens = [token for token in tokens if token.isalpha() and token not in stop_words]
    return tokens

df['tokens'] = df['value'].apply(preprocess_text)
```

	number	value	score	tokens
0	10.0	أقام المغرب مدارس جديده لتعليم الفتيات	2	[أقام, المغرب, مدارس, جديدة, لتعليم, الفتيات]
1	10.0	المغرب باشر إصلاحات اجتماعية لتحسين الوضع الاق	1	المغرب, باشر, إصلاحات, اجتماعية, لتحسين, الوض]
2	10.0	المغرب لم يقم بأي تغييرات اجتماعية يستحق الذكر	0	المغرب, لم, يقم, بأي, تغييرات, اجتماعية, يستح]
3	10.0	قام المغرب بإنشاء مستشفيات لتحسين الوضع الصحي	2	,قام, المغرب, بإنشاء, مستشفيات, لتحسين, الوضع]
4	10.0	فتح المغرب بعض المدارس لتعليم العلوم الحديثة	1	فتح, المغرب, بعض, المدارس, لتعليم, العلوم, ال]
05	Maki		0	f 14.1 + 1.1 1 41

Une étape essentielle dans le prétraitement des données textuelles est d'appliquer des transformations pour rendre le texte plus adapté à l'analyse. La fonction 'preprocess_text()' a été développée pour effectuer une série de prétraitements sur le texte (Tokenisation)

Une autre étape clé dans le prétraitement des données textuelles consiste à représenter les mots sous forme de vecteurs numériques, afin de pouvoir les utiliser comme entrées dans les algorithmes d'apprentissage automatique. La fonction 'get_feature_vector()' a été développée pour obtenir les vecteurs de caractéristiques correspondants à une liste de tokens donnée.

Cette fonction prend en entrée une liste de tokens, représentant les mots prétraités d'un texte, ainsi qu'un modèle de vectorisation des mots préalablement entraîné (un modèle Word2Vec). Elle initialise un vecteur de caractéristiques de dimension fixe (de taille 100) avec des valeurs nulles.

Ensuite, pour chaque token dans la liste, la fonction vérifie si le modèle de vectorisation contient une représentation vectorielle correspondante pour ce token. Si c'est le cas, le vecteur de caractéristiques est mis à jour en ajoutant le vecteur du token au vecteur existant.

Finalement, la fonction renvoie le vecteur de caractéristiques obtenu, qui représente la somme des vecteurs des tokens présents dans le texte d'origine.

la modélisation des données et Evaluation :

> modélisation des données :

```
train_data = df.sample(frac=0.80, random_state=42)
test_data = df.drop(train_data.index)

X_train = np.array(train_data['feature_vector'].tolist())
Y_train = np.array(train_data['score'])

X_test = np.array(test_data['feature_vector'].tolist())
Y_test = np.array(test_data['score'])
```

nous effectuons une division des données en un ensemble d'entraînement (train_data) et un ensemble de test (test_data) pour évaluer les performances de notre modèle.

```
from sklearn.linear_model import LogisticRegression
from sklearn.model_selection import train_test_split
from sklearn.neighbors import KNeighborsClassifier
from sklearn.tree import DecisionTreeClassifier
from sklearn.neural_network import MLPClassifier
from sklearn.naive_bayes import GaussianNB
from sklearn.svm import SVC
# KNN
def KNN():
    KNN = KNeighborsClassifier(n_neighbors=5)
    KNN.fit(X_train, Y_train)
    return KNN
# Decision Tree
def DT():
         DecisionTreeClassifier()
    DT.fit(X_train, Y_train)
    return DT
# ANN
def ANN():
    ANN = MLPClassifier(hidden_layer_sizes=(10, 10, 10), max_iter=1000)
    ANN.fit(X_train, Y_train)
    return ANN
# NB
def NB():
    NB = GaussianNB()
    NB.fit(X_train, Y_train)
    return NB
# SVM
def SVM_RBF():
    SVM = SVC(kernel='rbf',probability=True)
    SVM.fit(X_train, Y_train)
    return SVM
def SVM_LINEAR():
     SVM = SVC(kernel='linear',probability=True)
    SVM.fit(X_train, Y_train)
def SVM_POLY():
    SVM = SVC(kernel='poly',probability=True)
    SVM.fit(X_train, Y_train)
  return SVM
```

nous définissons plusieurs fonctions qui correspondent à différents algorithmes de classification que nous allons utiliser pour entraîner des modèles sur nos données.

Nous appelons chaque fonction correspondant à un algorithme de classification (KNN, DT, NB, ANN, SVM_RBF, SVM_LINEAR, SVM_POLY) pour obtenir les modèles entraînés respectifs.

Remarque : ces étapes sont répétées pour chaque question c a d on aura 10 modèles

- > Evaluation des 10 modèles :
 - modèle 1 (question 1):

```
------Classification Accuracy-----
 print("Accuracy KNN : {:.2f}%".format(accuracy_score(Y_test, y_pred_KNN)*100))
print("Accuracy KNN : {:.2f}%".format(accuracy_score(Y_test, y_pred_KNN)*100))
print("Accuracy DT : {:.2f}%".format(accuracy_score(Y_test, y_pred_DT)*100))
print("Accuracy ANN : {:.2f}%".format(accuracy_score(Y_test, y_pred_ANN)*100))
print("Accuracy NB : {:.2f}%".format(accuracy_score(Y_test, y_pred_NB)*100))
print("Accuracy SVM_RBF : {:.2f}%".format(accuracy_score(Y_test, y_pred_SVM_RBF)*100))
print("Accuracy SVM_LINEAR : {:.2f}%".format(accuracy_score(Y_test, y_pred_SVM_LINEAR)*100))
print("Accuracy SVM_POLY : {:.2f}%".format(accuracy_score(Y_test, y_pred_SVM_POLY)*100))
 Accuracy KNN: 94.44%
 Accuracy DT : 83.33%
 Accuracy ANN: 83.33%
Accuracy NB : 72.22%
Accuracy SVM_RBF : 83.33%
 Accuracy SVM_LINEAR : 83.33%
 Accuracy SVM_POLY : 88.89%
# Affichage du score AUC-ROC
print('AUC-ROC KNN score:', roc_auc_score(Y_test, y_scores_KNN, multi_class='ovr'))
print('AUC-ROC DT score:', roc_auc_score(Y_test, y_scores_DT, multi_class='ovr'))
print('AUC-ROC ANN score:', roc_auc_score(Y_test, y_scores_ANN, multi_class='ovr'))
print('AUC-ROC ANN score:', roc_auc_score(Y_test, y_scores_ANN, multi_class='ovr'))
print('AUC-ROC NB score:', roc_auc_score(Y_test, y_scores_NB, multi_class='ovr'))
print('AUC-ROC SVM_RBF score:', roc_auc_score(Y_test, y_scores_SVM_RBF, multi_class='ovr'))
print('AUC-ROC SVM_LINEAR score:', roc_auc_score(Y_test, y_scores_SVM_LINEAR, multi_class='ovr'))
print('AUC-ROC SVM_POLY score:', roc_auc_score(Y_test, y_scores_SVM_POLY, multi_class='ovr'))
AUC-ROC KNN score: 0.986441798941799
AUC-ROC DT score: 0.8672619047619047
AUC-ROC ANN score: 0.9100529100529101
AUC-ROC NB score: 0.8998677248677248
AUC-ROC SVM_RBF score: 0.9412037037037037
AUC-ROC SVM LINEAR score: 0.8971560846560847
AUC-ROC SVM_POLY score: 0.9375
```

```
print("Mean score RBF: ", cross_val_score(SVM_RBF, X_train, Y_train, cv=5).mean())
print("Mean score SVM LINEAR: ", cross_val_score(SVM_LINEAR, X_train, Y_train, cv=5).mean())
print("Mean score SVM POLY: ", cross_val_score(SVM_POLY, X_train, Y_train, cv=5).mean())
print("Mean score KNN: ", cross_val_score(KNN, X_train, Y_train, cv=5).mean())
#print("Mean score DT: ", cross_val_score(DT, X_train, Y_train, cv=5).mean)
print("Mean score ANN: ", cross_val_score(ANN, X_train, Y_train, cv=5).mean())
print("Mean score NB : ", cross_val_score(NB, X_train, Y_train, cv=5).mean())
Mean score RBF: 0.8028571428571428
Mean score SVM LINEAR: 0.6904761904761905
Mean score SVM POLY: 0.6171428571428571
Mean score KNN: 0.7876190476190476
             modèle 2 (question 2) :
                           ---Classification Accuracy-----
print("Accuracy KNN : {:.2f}%".format(accuracy_score(Y_test, y_pred_KNN)*100))
print("Accuracy DT : {:.2f}%".format(accuracy_score(Y_test, y_pred_DT)*100))
print("Accuracy NB : {:.2f}%".format(accuracy_score(Y_test, y_pred_NB)*100))
print("Accuracy SVM_RBF : {:.2f}%".format(accuracy_score(Y_test, y_pred_SVM_RBF)*100))
print("Accuracy SVM_LINEAR : {:.2f}%".format(accuracy_score(Y_test, y_pred_SVM_LINEAR)*100))
print("Accuracy SVM_POLY : {:.2f}%".format(accuracy_score(Y_test, y_pred_SVM_POLY)*100))
print("Accuracy ANN : {:.2f}%".format(accuracy_score(Y_test, y_pred_ANN)*100))
Accuracy KNN: 89.47%
Accuracy DT : 63.16%
Accuracy NB: 89.47%
Accuracy SVM_RBF : 89.47%
Accuracy SVM_LINEAR : 57.89%
Accuracy SVM POLY : 57.89%
y_scores_KNN = KNN.predict_proba(X_test)
y_scores_DT = DT.predict_proba(X_test)
y_scores_ANN = ANN.predict_proba(X_test)
y_scores_NB = NB.predict_proba(X_test)
y_scores_SVM_RBF = SVM_RBF.predict_proba(X_test)
y_scores_SVM_LINEAR = SVM_LINEAR.predict_proba(X_test)
y_scores_SVM_POLY = SVM_POLY.predict_proba(X_test)
# Affichage du score AUC-ROC
print('AUC-ROC KNN score:', roc_auc_score(Y_test, y_scores_KNN, multi_class='ovr'))
```

y_scores_ANN = ANN.predict_proba(X_test)
y_scores_NB = NB.predict_proba(X_test)
y_scores_SVM_RBF = SVM_RBF.predict_proba(X_test)
y_scores_SVM_LINEAR = SVM_LINEAR.predict_proba(X_test)
y_scores_SVM_LINEAR = SVM_DOLY.predict_proba(X_test)
Affichage du score AUC-ROC
print('AUC-ROC KNN score:', roc_auc_score(Y_test, y_scores_KNN, multi_class='ovr'))
print('AUC-ROC DT score:', roc_auc_score(Y_test, y_scores_DT, multi_class='ovr'))
print('AUC-ROC NB score:', roc_auc_score(Y_test, y_scores_NB, multi_class='ovr'))
print('AUC-ROC SVM_RBF score:', roc_auc_score(Y_test, y_scores_SVM_RBF, multi_class='ovr'))
print('AUC-ROC SVM_LINEAR score:', roc_auc_score(Y_test, y_scores_SVM_LINEAR, multi_class='ovr'))
print('AUC-ROC SVM_POLY score:', roc_auc_score(Y_test, y_scores_SVM_POLY, multi_class='ovr'))
print('AUC-ROC ANN score:', roc_auc_score(Y_test, y_scores_ANN, multi_class='ovr'))

AUC-ROC SVM_RBF score: 0.99065656565657

AUC-ROC SVM_LINEAR score: 0.78383838383839

AUC-ROC SVM_POLY score: 0.765151515151515

print("Mean score RBF: ", cross_val_score(SVM_RBF, X_train, Y_train, cv=5).mean())

print("Mean score ANN: ", cross_val_score(ANN, X_train, Y_train, cv=5).mean())

print("Mean score NB : ", cross_val_score(NB, X_train, Y_train, cv=5).mean())

print("Mean score SVM LINEAR: ", cross_val_score(SVM_LINEAR, X_train, Y_train, cv=5).mean())

print("Mean score KNN: ", cross_val_score(KNN, X_train, Y_train, cv=5).mean())

print("Mean score SVM POLY: ", cross_val_score(SVM_POLY, X_train, Y_train, cv=5).mean())

print("Mean score DT: ", cross_val_score(DT, X_train, Y_train, cv=5).mean)

AUC-ROC DT score: 0.7236652236652237 AUC-ROC NB score: 0.9813131313131312

• modèle 3 (question 3):

```
-Classification Accuracy-
print("Accuracy KNN : {:.2f}%".format(accuracy_score(Y_test, y_pred_KNN)*100))
print("Accuracy DT : {:.2f}%".format(accuracy_score(Y_test, y_pred_DT)*100))
print("Accuracy ANN : {:.2f}%".format(accuracy_score(Y_test, y_pred_ANN)*100))
print("Accuracy NB : {:.2f}%".format(accuracy_score(Y_test, y_pred_NB)*100))
print("Accuracy SVM_RBF : {:.2f}%".format(accuracy_score(Y_test, y_pred_SVM_RBF)*100))
print("Accuracy SVM_LINEAR : {:.2f}%".format(accuracy_score(Y_test, y_pred_SVM_LINEAR)*100))
print("Accuracy SVM_POLY : {:.2f}%".format(accuracy_score(Y_test, y_pred_SVM_POLY)*100))
Accuracy KNN : 77.78%
Accuracy DT : 66.67%
Accuracy ANN: 94.44%
Accuracy NB : 83.33%
Accuracy SVM_RBF : 94.44%
Accuracy SVM_LINEAR : 88.89%
Accuracy SVM_POLY : 55.56%
print("Mean score RBF: ", cross_val_score(SVM_RBF, X_train, Y_train, cv=5).mean())
print("Mean score SVM LINEAR: ", cross_val_score(SVM_LINEAR, X_train, Y_train, cv=5).mean())
print("Mean score SVM POLY: ", cross_val_score(SVM_POLY, X_train, Y_train, cv=5).mean())
print("Mean score ANN: ", cross_val_score(ANN, X_train, Y_train, cv=) print("Mean score NB: ", cross_val_score(ANN, X_train, Y_train, cv=5).mean()) print("Mean score KNN: ", cross_val_score(KNN, X_train, Y_train, cv=5).mean()) print("Mean score DT: ", cross_val_score(DT, X_train, Y_train, cv=5).mean())
Mean score RBF: 0.8142857142857144
Mean score SVM LINEAR: 0.73333333333333333
Mean score SVM POLY: 0.6761904761904762
Mean score ANN: 0.9285714285714285
Mean score NB : 0.7723809523809524
Mean score KNN: 0.818095238095238
y_scores_KNN = KNN.predict_proba(X_test)
y_scores_DT = DT.predict_proba(X_test)
y_scores_ANN = ANN.predict_proba(X_test)
y_scores_NB = NB.predict_proba(X_test)
y_scores_SVM_RBF = SVM_RBF.predict_proba(X_test)
y_scores_SVM_LINEAR = SVM_LINEAR.predict_proba(X_test)
y_scores_SVM_POLY = SVM_POLY.predict_proba(X_test)
# Affichage du score AUC-ROC
print('AUC-ROC KNN score:', roc_auc_score(Y_test, y_scores_KNN, multi_class='ovr'))
print('AUC-ROC DT score:', roc_auc_score(Y_test, y_scores_DT, multi_class='ovr'))
print('AUC-ROC NB score:', roc_auc_score(Y_test, y_scores_NB, multi_class='ovr'))
print('AUC-ROC SVM_RBF score:', roc_auc_score(Y_test, y_scores_SVM_RBF, multi_class='ovr'))
print('AUC-ROC SVM_LINEAR score:', roc_auc_score(Y_test, y_scores_SVM_LINEAR, multi_class='ovr'))
print('AUC-ROC SVM_POLY score:', roc_auc_score(Y_test, y_scores_SVM_POLY, multi_class='ovr'))
print('AUC-ROC ANN score:', roc_auc_score(Y_test, y_scores_ANN, multi_class='ovr'))
AUC-ROC KNN score: 0.8913690476190476
AUC-ROC DT score: 0.7779761904761905
AUC-ROC NB score: 0.9696428571428571
AUC-ROC SVM_RBF score: 0.9898809523809525
AUC-ROC SVM_LINEAR score: 0.9898809523809525
AUC-ROC SVM_POLY score: 0.6964285714285715
AUC-ROC ANN score: 1.0
```

• modèle 4 (question 4):

```
## -----:
print("Accuracy KNN : {:.2f}%".format(accuracy_score(Y_test, y_pred_KNN)*100))
print("Accuracy DT : {:.2f}%".format(accuracy_score(Y_test, y_pred_DT)*100))
print("Accuracy ANN : {:.2f}%".format(accuracy_score(Y_test, y_pred_ANN)*100))
print("Accuracy NB : {:.2f}%".format(accuracy_score(Y_test, y_pred_NB)*100))
print("Accuracy SVM_RBF : {:.2f}%".format(accuracy_score(Y_test, y_pred_SVM_RBF)*100))
print("Accuracy SVM_LINEAR : {:.2f}%".format(accuracy_score(Y_test, y_pred_SVM_LINEAR)*100))
print("Accuracy SVM_POLY : {:.2f}%".format(accuracy_score(Y_test, y_pred_SVM_POLY)*100))
Accuracy KNN: 62.96%
Accuracy DT : 51.85%
Accuracy ANN : 48.15%
Accuracy NB : 81.48%
Accuracy SVM_RBF : 85.19%
Accuracy SVM_LINEAR : 29.63%
Accuracy SVM_POLY : 66.67%
y_scores_KNN = KNN.predict_proba(X_test)
y_scores_DT = DT.predict_proba(X_test)
y_scores_ANN = ANN.predict_proba(X_test)
y_scores_NB = NB.predict_proba(X_test)
y_scores_SVM_RBF = SVM_RBF.predict_proba(X_test)
y_scores_SVM_LINEAR = SVM_LINEAR.predict_proba(X_test)
y_scores_SVM_POLY = SVM_POLY.predict_proba(X_test)
# Affichage du score AUC-ROC
print('AUC-ROC KNN score:', roc_auc_score(Y_test, y_scores_KNN, multi_class='ovr'))
print('AUC-ROC DT score:', roc_auc_score(Y_test, y_scores_DT, multi_class='ovr'))
print('AUC-ROC ANN score:', roc_auc_score(Y_test, y_scores_ANN, multi_class='ovr'))
print('AUC-ROC NB score:', roc_auc_score(Y_test, y_scores_NB, multi_class='ovr'))
print('AUC-ROC SVM_RBF score:', roc_auc_score(Y_test, y_scores_SVM_RBF, multi_class='ovr'))
print('AUC-ROC SVM_LINEAR score:', roc_auc_score(Y_test, y_scores_SVM_LINEAR, multi_class='ovr'))
print('AUC-ROC SVM_POLY score:', roc_auc_score(Y_test, y_scores_SVM_POLY, multi_class='ovr'))
AUC-ROC KNN score: 0.7867864104706209
AUC-ROC DT score: 0.6257414369256474
AUC-ROC ANN score: 0.7585700362016152
AUC-ROC NB score: 0.8995265942634364
AUC-ROC SVM_RBF score: 0.9328947368421052
AUC-ROC SVM_LINEAR score: 0.8546365914786967
AUC-ROC SVM_POLY score: 0.817000835421888
print("Mean score RBF: ", cross_val_score(SVM_RBF, X_train, Y_train, cv=5).mean())
print("Mean score SVM LINEAR: ", cross_val_score(SVM_LINEAR, X_train, Y_train, cv=5).mean())
print("Mean score SVM POLY: ", cross_val_score(SVM_POLY, X_train, Y_train, cv=5).mean())
print("Mean score ANN: ", cross_val_score(ANN, X_train, Y_train, cv=5).mean())
print("Mean score NB: ", cross_val_score(NB, X_train, Y_train, cv=5).mean())
print("Mean score KNN: ", cross_val_score(KNN, X_train, Y_train, cv=5).mean())
#print("Mean score DT: ", cross_val_score(DT, X_train, Y_train, cv=5).mean)
Mean score RBF: 0.6871794871794872
Mean score SVM LINEAR: 0.358974358974359
Mean score SVM POLY: 0.5141025641025642
Mean score ANN: 0.6269230769230769
```

Mean score NB : 0.5923076923076923 Mean score KNN: 0.5282051282051282

• modèle 5 (question):

```
-----Classification Accuracy----
print("Accuracy KNN : {:.2f}%".format(accuracy_score(Y_test, y_pred_KNN)*100))
print("Accuracy DT : {:.2f}%".format(accuracy_score(Y_test, y_pred_DT)*100))
print("Accuracy ANN : {:.2f}%".format(accuracy_score(Y_test, y_pred_ANN)*100))
print( Accuracy Ann : { ..2f}%".format(accuracy_score(Y_test, y_pred_NB)*100))
print("Accuracy SVM_RBF : { :.2f}%".format(accuracy_score(Y_test, y_pred_SVM_RBF)*100))
print("Accuracy SVM_LINEAR : { :.2f}%".format(accuracy_score(Y_test, y_pred_SVM_LINEAR)*100))
print("Accuracy SVM_LINEAR : { :.2f}%".format(accuracy_score(Y_test, y_pred_SVM_LINEAR)*100))
print("Accuracy SVM_POLY : {:.2f}%".format(accuracy_score(Y_test, y_pred_SVM_POLY)*100))
Accuracy KNN : 70.00%
Accuracy DT : 55.00%
Accuracy ANN: 80.00%
Accuracy NB : 65.00%
Accuracy SVM_RBF : 65.00%
Accuracy SVM_LINEAR : 55.00%
Accuracy SVM_POLY : 70.00%
```

```
print("Mean score SVM POLY: ", cross_val_score(SVM_POLY, X_train, Y_train, cv=5).mean())
print("Mean score ANN: ", cross_val_score(ANN, X_train, Y_train, cv=5).mean())
print("Mean score NB: ", cross_val_score(NB, X_train, Y_train, cv=5).mean())
print("Mean score KNN: ", cross_val_score(KNN, X_train, Y_train, cv=5).mean())
print("Mean score RBF: ", cross_val_score(SVM_RBF, X_train, Y_train, cv=5).mean())
#print("Mean score DT: ", cross_val_score(DT, X_train, Y_train, cv=5).mean)
print("Mean score SVM_LINEAR: ", cross_val_score(SVM_LINEAR, X_train, Y_train, cv=5).mean())

Mean score SVM POLY: 0.558102766798419
Mean score ANN: 0.6632411067193675
Mean score NB: 0.6640316205533596
Mean score KNN: 0.6209486166007905
Mean score RBF: 0.6810276679841897
Mean score SVM_LINEAR: 0.574703557312253
```

• modèle 6 (question 6):

```
## -----Classification Accuracy-----
print("Accuracy KNN : {:.2f}%".format(accuracy_score(Y_test, y_pred_KNN)*100))
print("Accuracy DT : {:.2f}%".format(accuracy_score(Y_test, y_pred_DT)*100))
print("Accuracy ANN : {:.2f}%".format(accuracy_score(Y_test, y_pred_ANN)*100))
print("Accuracy NB : {:.2f}%".format(accuracy_score(Y_test, y_pred_NB)*100))
print("Accuracy SVM_RBF : {:.2f}%".format(accuracy_score(Y_test, y_pred_SVM_RBF)*100))
print("Accuracy SVM_LINEAR : {:.2f}%".format(accuracy_score(Y_test, y_pred_SVM_LINEAR)*100))
print("Accuracy SVM_POLY : {:.2f}%".format(accuracy_score(Y_test, y_pred_SVM_POLY)*100))
Accuracy KNN: 90.62%
Accuracy DT : 71.88%
Accuracy ANN: 81.25%
Accuracy NB : 87.50%
Accuracy SVM_RBF : 78.12%
Accuracy SVM_LINEAR : 50.00%
Accuracy SVM_POLY : 50.00%
print("Mean score SVM LINEAR: ", cross_val_score(SVM_LINEAR, X_train, Y_train, cv=5).mean())
print("Mean score SVM LINEAR: ", cross_val_score(SVM_LINEAR, X_train, Y_train, cv=5).mean
print("Mean score SVM POLY: ", cross_val_score(SVM_POLY, X_train, Y_train, cv=5).mean())
print("Mean score ANN: ", cross_val_score(ANN, X_train, Y_train, cv=5).mean())
print("Mean score NB : ", cross_val_score(NB, X_train, Y_train, cv=5).mean())
print("Mean score RBF: ", cross_val_score(SVM_RBF, X_train, Y_train, cv=5).mean())
print("Mean score KNN: ", cross_val_score(KNN, X_train, Y_train, cv=5).mean())
#print("Mean score DT: ", cross_val_score(DT, X_train, Y_train, cv=5).mean)
Mean score SVM LINEAR: 0.69333333333333334
Mean score SVM POLY: 0.679999999999999
Mean score ANN: 0.906666666666666
Mean score NB : 0.906666666666666
Mean score RBF: 0.7866666666666667
Mean score KNN: 0.906666666666666
# Affichage du score AUC-ROC
print('AUC-ROC KNN score:', roc_auc_score(Y_test, y_scores_KNN, multi_class='ovr'))
print('AUC-ROC DT score:', roc_auc_score(Y_test, y_scores_DT, multi_class='ovr'))
print('AUC-ROC ANN score:', roc_auc_score(Y_test, y_scores_ANN, multi_class='ovr'))
print('AUC-ROC ANN score:', roc_auc_score(Y_test, y_scores_ANN, multi_class='ovr'))
print('AUC-ROC NB score:', roc_auc_score(Y_test, y_scores_NB, multi_class='ovr'))
print('AUC-ROC SVM_RBF score:', roc_auc_score(Y_test, y_scores_SVM_RBF, multi_class='ovr'))
print('AUC-ROC SVM_LINEAR score:', roc_auc_score(Y_test, y_scores_SVM_LINEAR, multi_class='ovr'))
print('AUC-ROC SVM_POLY score:', roc_auc_score(Y_test, y_scores_SVM_POLY, multi_class='ovr'))
AUC-ROC KNN score: 0.9617075517075517
AUC-ROC DT score: 0.8034343434343435
AUC-ROC ANN score: 0.9896151996151996
AUC-ROC NB score: 0.9742568542568543
AUC-ROC SVM RBF score: 0.9894949494949495
AUC-ROC SVM_LINEAR score: 0.9832611832611833
AUC-ROC SVM POLY score: 0.9606060606060606
```

modèle 7 (question 7) :

```
print("Accuracy KNN : {:.2f}%".format(accuracy_score(Y_test, y_pred_KNN)*100))
print("Accuracy DT : {:.2f}%".format(accuracy_score(Y_test, y_pred_DT)*100))
print("Accuracy ANN : {:.2f}%".format(accuracy_score(Y_test, y_pred_ANN)*100))
print( Accuracy ANN : {:.2f}% .Tormat(accuracy_score('_test, y_pred_ANN)*100))
print("Accuracy NB : {:.2f}%".format(accuracy_score(Y_test, y_pred_NB)*100))
print("Accuracy SVM_RBF : {:.2f}%".format(accuracy_score(Y_test, y_pred_SVM_RBF)*100))
print("Accuracy SVM_LINEAR : {:.2f}%".format(accuracy_score(Y_test, y_pred_SVM_LINEAR)*100))
print("Accuracy SVM_POLY : {:.2f}%".format(accuracy_score(Y_test, y_pred_SVM_POLY)*100))
Accuracy KNN: 87.50%
Accuracy DT : 81.25%
Accuracy ANN : 93.75%
Accuracy NB : 87.50%
Accuracy SVM_RBF : 87.50%
Accuracy SVM_LINEAR : 50.00%
Accuracy SVM_POLY : 50.00%
 print("Mean score RBF: ", cross_val_score(SVM_RBF, X_train, Y_train, cv=5).mean())
print("Mean score KBP.", cross_val_score(SVM_LINEAR, X_train, Y_train, cv=5).mean())
print("Mean score SVM POLY: ", cross_val_score(SVM_POLY, X_train, Y_train, cv=5).mean())
print("Mean score ANN: ", cross_val_score(ANN, X_train, Y_train, cv=5).mean())
print("Mean score NB: ", cross_val_score(NB, X_train, Y_train, cv=5).mean())
print("Mean score KNN: ", cross_val_score(KNN, X_train, Y_train, cv=5).mean())
 print("Mean score DT: ", cross_val_score(DT, X_train, Y_train, cv=5).mean)
Mean score RBF: 0.9064102564102564
 Mean score SVM LINEAR: 0.6205128205128204
 Mean score SVM POLY: 0.6358974358974359
 Mean score ANN: 0.8910256410256409
Mean score NB : 0.8589743589743591
Mean score KNN: 0.7448717948717949
# Affichage du score AUC-ROC
print('AUC-ROC KNN score:', roc_auc_score(Y_test, y_scores_KNN, multi_class='ovr'))
print('AUC-ROC DT score:', roc_auc_score(Y_test, y_scores_DT, multi_class='ovr'))
print('AUC-ROC ANN score:', roc_auc_score(Y_test, y_scores_ANN, multi_class='ovr'))
print('AUC-ROC NB score:', roc_auc_score(Y_test, y_scores_NB, multi_class='ovr'))
 print('AUC-ROC SVM_RBF score:', roc_auc_score(Y_test, y_scores_SVM_RBF, multi_class='ovr'))
 print('AUC-ROC SVM_LINEAR score:', roc_auc_score(Y_test, y_scores_SVM_LINEAR, multi_class='ovr'))
 print('AUC-ROC SVM_POLY score:', roc_auc_score(Y_test, y_scores_SVM_POLY, multi_class='ovr'))
 AUC-ROC KNN score: 0.9453125
 AUC-ROC DT score: 0.84027777777778
 AUC-ROC ANN score: 1.0
 AUC-ROC NB score: 1.0
 AUC-ROC SVM_RBF score: 1.0
 AUC-ROC SVM_LINEAR score: 0.7100694444444445
 AUC-ROC SVM_POLY score: 0.6493055555555555
```

• modèle 8 (question 8):

```
# Affichage du score AUC-ROC
print('AUC-ROC KNN score:', roc_auc_score(Y_test, y_scores_KNN, multi_class='ovr'))
print('AUC-ROC DT score:', roc_auc_score(Y_test, y_scores_DT, multi_class='ovr'))
print('AUC-ROC ANN score:', roc_auc_score(Y_test, y_scores_ANN, multi_class='ovr'))
print('AUC-ROC NB score:', roc_auc_score(Y_test, y_scores_NB, multi_class='ovr'))
print('AUC-ROC SVM_RBF score:', roc_auc_score(Y_test, y_scores_SVM_RBF, multi_class='ovr'))
print('AUC-ROC SVM_LINEAR score:', roc_auc_score(Y_test, y_scores_SVM_LINEAR, multi_class='ovr'))
print('AUC-ROC SVM_POLY score:', roc_auc_score(Y_test, y_scores_SVM_POLY, multi_class='ovr'))
AUC-ROC KNN score: 0.9529848171152518
AUC-ROC DT score: 0.7289222452265931
AUC-ROC ANN score: 0.9361858753163101
AUC-ROC NB score: 0.9224522659305268
AUC-ROC SVM_RBF score: 0.9108005521049
AUC-ROC SVM_LINEAR score: 0.8420519898780768
AUC-ROC SVM POLY score: 0.9001150218541523
print("Mean score RBF: ", cross_val_score(SVM_RBF, X_train, Y_train, cv=5).mean())
print("Mean score SVM LINEAR: ", cross_val_score(SVM_LINEAR, X_train, Y_train, cv=5).mean())
print("Mean score SVM POLY: ", cross_val_score(SVM_POLY, X_train, Y_train, cv=5).mean())
print("Mean score KNN: ", cross_val_score(KNN, X_train, Y_train, cv=5).mean())
#print("Mean score DT: ", cross_val_score(DT, X_train, Y_train, cv=5).mean)
print("Mean score ANN: ", cross_val_score(ANN, X_train, Y_train, cv=5).mean())
print("Mean score NB : ", cross_val_score(NB, X_train, Y_train, cv=5).mean())
Mean score RBF: 0.69583333333333333
Mean score SVM LINEAR: 0.6291666666666667
Mean score SVM POLY: 0.655
Mean score KNN: 0.7575000000000001
Mean score ANN: 0.7575
Mean score NB: 0.6824999999999999
               modèle 9 (question 9) :
                           -Classification Accuracy--
Accuracy KNN : 83.33%
Accuracy DT : 55.56%
Accuracy ANN: 83.33%
Accuracy NB: 83.33%
Accuracy SVM_RBF: 77.78%
Accuracy SVM_LINEAR: 38.89%
Accuracy SVM_POLY : 50.00%
# Affichage du score AUC-ROC
print('AUC-ROC KNN score:', roc_auc_score(Y_test, y_scores_KNN, multi_class='ovr'))
print('AUC-ROC DT score:', roc_auc_score(Y_test, y_scores_DT, multi_class='ovr'))
print('AUC-ROC ANN score:', roc_auc_score(Y_test, y_scores_ANN, multi_class='ovr'))
print('AUC-ROC NB score:', roc_auc_score(Y_test, y_scores_NB, multi_class='ovr'))
print('AUC-ROC SVM_RBF score:', roc_auc_score(Y_test, y_scores_SVM_RBF, multi_class='ovr'))
print('AUC-ROC SVM_LINEAR score:', roc_auc_score(Y_test, y_scores_SVM_LINEAR, multi_class='ovr'))
print('AUC-ROC SVM_POLY score:', roc_auc_score(Y_test, y_scores_SVM_POLY, multi_class='ovr'))
AUC-ROC KNN score: 0.9807900432900433
AUC-ROC DT score: 0.6720779220779219
AUC-ROC ANN score: 0.9761904761904762
AUC-ROC NB score: 0.9512987012987013
AUC-ROC SVM_RBF score: 0.9512987012987013
AUC-ROC SVM_LINEAR score: 0.9307359307359307
```

AUC-ROC SVM_POLY score: 0.7521645021645021

• modèle 10 (question 10):

```
-----Classification Accuracy----
 print("Accuracy KNN : {:.2f}%".format(accuracy_score(Y_test, y_pred_KNN)*100))
print("Accuracy KNN : {:.2f}%".format(accuracy_score(Y_test, y_pred_KNN)*100))
print("Accuracy NB : {:.2f}%".format(accuracy_score(Y_test, y_pred_NB)*100))
print("Accuracy SVM_RBF : {:.2f}%".format(accuracy_score(Y_test, y_pred_SVM_RBF)*100))
print("Accuracy SVM_LINEAR : {:.2f}%".format(accuracy_score(Y_test, y_pred_SVM_LINEAR)*100))
print("Accuracy SVM_POLY : {:.2f}%".format(accuracy_score(Y_test, y_pred_SVM_POLY)*100))
print("Accuracy ANN : {:.2f}%".format(accuracy_score(Y_test, y_pred_ANN)*100))
print("Accuracy DT : {:.2f}%".format(accuracy_score(Y_test, y_pred_DT)*100))
 Accuracy KNN: 73.33%
 Accuracy NB : 53.33%
Accuracy SVM_RBF : 66.67%
Accuracy SVM_LINEAR : 60.00%
 Accuracy SVM_POLY : 60.00%
 Accuracy ANN: 73.33%
print("Mean score RBF: ", cross_val_score(SVM_RBF, X_train, Y_train, cv=5).mean())
print("Mean score RBF: ", cross_val_score(SVM_RBF, X_train, Y_train, cv=5).mean())
print("Mean score SVM LINEAR: ", cross_val_score(SVM_LINEAR, X_train, Y_train, cv=5).mean())
print("Mean score SVM POLY: ", cross_val_score(SVM_POLY, X_train, Y_train, v=5).mean())
print("Mean score ANN: ", cross_val_score(ANN, X_train, Y_train, cv=5).mean())
print("Mean score NB: ", cross_val_score(NB, X_train, Y_train, cv=5).mean())
print("Mean score KNN: ", cross_val_score(KNN, X_train, Y_train, cv=5).mean())
Mean score RBF: 0.6588235294117647
Mean score SVM LINEAR: 0.3647058823529412
Mean score SVM POLY: 0.6117647058823529
Mean score ANN: 0.6
Mean score NB : 0.611764705882353
Mean score KNN: 0.5529411764705883
 y_scores_KNN = KNN.predict_proba(X_test)
y_scores_DT = DT.predict_proba(X_test)
y_scores_ANN = ANN.predict_proba(X_test)
y_scores_NB = NB.predict_proba(X_test)
y_scores_SVM_RBF = SVM_RBF.predict_proba(X_test)
y_scores_SVM_LINEAR = SVM_LINEAR.predict_proba(X_test)
y_scores_SVM_POLY = SVM_POLY.predict_proba(X_test)
# Affichage du score AUC-ROC
# Affichage du score AUC-ROC
print('AUC-ROC KNN score:', roc_auc_score(Y_test, y_scores_KNN, multi_class='ovr'))
print('AUC-ROC DT score:', roc_auc_score(Y_test, y_scores_DT, multi_class='ovr'))
print('AUC-ROC ANN score:', roc_auc_score(Y_test, y_scores_ANN, multi_class='ovr'))
print('AUC-ROC NB score:', roc_auc_score(Y_test, y_scores_NB, multi_class='ovr'))
print('AUC-ROC SVM_RBF score:', roc_auc_score(Y_test, y_scores_SVM_RBF, multi_class='ovr'))
print('AUC-ROC SVM_LINEAR score:', roc_auc_score(Y_test, y_scores_SVM_LINEAR, multi_class='ovr'))
print('AUC-ROC SVM_POLY score:', roc_auc_score(Y_test, y_scores_SVM_POLY, multi_class='ovr'))
AUC-ROC KNN score: 0.8976301476301476
 AUC-ROC DT score: 0.8573556073556073
AUC-ROC ANN score: 0.8448804282137615
AUC-ROC NB score: 0.7557411724078391
AUC-ROC SVM_RBF score: 0.940796857463524
AUC-ROC SVM_LINEAR score: 0.1962358629025296
AUC-ROC SVM_POLY score: 0.8294051627384961
```

Resumé:

Modèle	Algorithme adéquat	Accuracy	Auc-Roc	Cross Validation
1	KNN	94%	0.98	79%
2	SVM_RBF	89%	0.99	92%
3	KNN	74	0.89	56%
4	Naive Bayes	82%	0.89	60%
5	ANN	80%	0.85	66%
6	KNN	90%	0.96	90%
7	ANN	93%	1	89%
8	SVM_RBF	89%	0.91	70%
9	Naive Bayes	83%	0.95	87%
10	KNN	74	0.89	56%

Remarque:

Parfois, il y a des modèles où la valeur de validation croisée n'est pas élevée, mais où l'AUC-ROC et l'Accuracy sont élevées. Cela ne pose pas le problème d'Overfitting (surapprentissage), car ces modèles sont bien testés sur de nouvelles données.

Enregistrement des modèles :

```
import pickle
model_path = 'C:/Users/HP/Desktop/save_models/model_question4.h5'
with open(model_path, 'wb') as file:
    pickle.dump((NB, model), file)
```

16/05/2023 23:08	Fichier H5	524 Ko
17/05/2023 14:43	Fichier H5	307 Ko
17/05/2023 14:49	Fichier H5	419 Ko
17/05/2023 14:49	Fichier H5	391 Ko
17/05/2023 14:50	Fichier H5	651 Ko
17/05/2023 14:50	Fichier H5	464 Ko
17/05/2023 14:52	Fichier H5	306 Ko
17/05/2023 14:53	Fichier H5	360 Ko
17/05/2023 14:53	Fichier H5	422 Ko
21/05/2023 17:38	Fichier H5	326 Ko
	17/05/2023 14:43 17/05/2023 14:49 17/05/2023 14:49 17/05/2023 14:50 17/05/2023 14:50 17/05/2023 14:52 17/05/2023 14:53	17/05/2023 14:43 Fichier H5 17/05/2023 14:49 Fichier H5 17/05/2023 14:49 Fichier H5 17/05/2023 14:50 Fichier H5 17/05/2023 14:50 Fichier H5 17/05/2023 14:52 Fichier H5 17/05/2023 14:53 Fichier H5 17/05/2023 14:53 Fichier H5

Déploiement du modèle :

Dans ce projet, on a utilisé FastAPI pour développer le backend de notre application et Angular pour le frontend. Nous avons choisi FastAPI en raison de sa performance élevée et de sa prise en charge intégrée de GraphQL, qui nous a permis de mettre en place une API GraphQL pour la communication entre le frontend et le backend.

> FastAPI:

```
from fastapi import FastAPI
from strawberry.asgi import GraphQL
import strawberry
from fastapi.middleware.cors import CORSMiddleware
from typing import List
import models.ready_model as ready_model

app = FastAPI()
```

L'importation de dix modules :

```
➤ models
> __pycache__
≡ model_question1.h5
≡ model_question2.h5
≡ model_question3.h5
≡ model_question4.h5
≡ model_question5.h5
≡ model_question6.h5
≡ model_question7.h5
≡ model_question9.h5
≡ model_question9.h5
≡ model_question10.h5
๗ ready_model.py
```

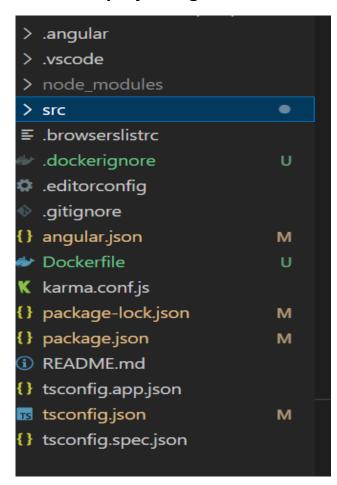
```
ready_model.py X
models > 🕏 ready_model.py > 🕥 predict
    import pickle
      import numpy as np
      from nltk.tokenize import word_tokenize
     import nltk
     nltk.download('punkt')
     import warnings
     warnings.filterwarnings('ignore')
 10 def predict(input_ans , num_ques):
          model_path = "model_question"+num_ques + ".h5"
          with open(model_path, 'rb') as file:
             model, model_word2vec = pickle.load(file)
         input_ans = preprocces_input(input_ans, model_word2vec)
 15
         input_ans = input_ans.reshape(1, -1)
          pred = model.predict(input_ans)
          result = pred[0]
          return result
```

La prédiction du score de chaque réponse et le retourner au front :

```
@strawberry.type
class Mutation:
    @strawberry.mutation
    def create_question(self, id: str, answer: str) -> List[Question]:
        if(id=="1"):
            questions.clear()
            score=ready_model.predict(answer, id)
            questions.append(Question(id=id, answer=answer, score=score))
            return questions
```

> Angular :

La structure du projet Angular :



Dans App on y crée quatre component : page principale, page des questions, page des réponses et le navbar :



> Graphql:

Pour faciliter la communication entre le frontend et le backend, on a utilisé GraphQL.

Intégration en fastApi (coté backend) :

On a utilisé Strawberry qui est une bibliothèque Python qui permet de créer des serveurs GraphQL de manière simple et intuitive. Elle offre une syntaxe déclarative pour définir des schémas GraphQL, des types de données, des requêtes et des résolveurs.

```
@strawberry.type
class Question:
   id: str
    answer: str
    score: int
questions: List[Question] = []
@strawberry.type
class Query:
   @strawberry.field
    def questions(self) -> List[Question]:
        return questions
@strawberry.type
class Mutation:
    @strawberry.mutation
    def create_question(self, id: str, answer: str) -> List[Question]:
        if(id=="1"):
            questions.clear()
        score=ready_model.predict(answer, id)
        questions.append(Question(id=id, answer=answer, score=score))
        return questions
```

```
schema = strawberry.Schema(query=Query, mutation=Mutation)

graphql_app = GraphQL(schema)
app.add_route("/graphql", graphql_app)
app.add_websocket_route("/graphql", graphql_app))
```

Intégration en Angular (Coté frontend) :

graphql.module:

Apollo facilite l'exécution de requêtes GraphQL, la gestion du cache, la souscription aux mises à jour en temps réel et bien plus encore. Il fournit également des outils et des composants pour faciliter l'intégration d'Apollo dans votre application Angular, ce qui vous permet de développer rapidement des fonctionnalités basées sur GraphQL.

```
import { APOLLO_OPTIONS, ApolloModule } from 'apollo-angular';
import { HttpLink } from 'apollo-angular/http';
import { NgModule } from '@angular/core';
import { ApolloClientOptions, InMemoryCache } from '@apollo/client/core';
const uri = 'http://127.0.0.1:8000/graphql'; // <-- add the URL of the GraphQL server here</pre>
export function createApollo(httpLink: HttpLink): ApolloClientOptions<any> {
   link: httpLink.create({ uri }),
   cache: new InMemoryCache(),
@NgModule({
 exports: [ApolloModule],
 providers: [
     provide: APOLLO_OPTIONS,
     useFactory: createApollo,
     deps: [HttpLink],
    },
export class GraphQLModule {}
```

Graphql.operations:

Les constantes GET_ANSWERS et CREATE_QUESTION contiennent les requêtes pour récupérer des réponses de questions existantes et créer une nouvelle question, respectivement.

Mise en conteneur avec Docker

> Containerisation du backend avec Docker :

```
✓ fastApi
> __pycache__
> .venv
> models
> venv

dockerignore
docker-compose.yml

✓ Dockerfile
main.py
requirements.txt
```

a) Construction de l'image Docker:

```
** Dockerfile ×

fastApi > ** Dockerfile > ◆ FROM

1    FROM python
2    WORKDIR /app
3    COPY ./requirements.txt .
4    COPY ./models .
5    RUN pip install -r requirements.txt
6    COPY .
7    EXPOSE 8000
```

b) Création du conteneur Docker :

```
(.venv) PS C:\Users\Dell LATITUDE\Desktop\docker\back\fastApi> docker-compose up
[+] Building 157.8s (11/11) FINISHED
=> => transferring dockerfile: 32B
=> [internal] load .dockerignore
=> => transferring context: 34B
=> [internal] load metadata for docker.io/library/python:latest
=> => transferring context: 1.31MB
=> CACHED [2/6] WORKDIR /app
=> CACHED [3/6] COPY ./requirements.txt .
=> [4/6] COPY ./models .
=> [5/6] RUN pip install -r requirements.txt
=> exporting to image
=> => writing image sha256:c8a80ed85c5d56746ab732ca99f4d28ed08b5bd76ad806c2a77b6e161d515ce7
Use 'docker scan' to run Snyk tests against images to find vulnerabilities and learn how to fix them
[+] Running 2/2
  - Network fastapi_default Created
 - Container fastapi-app-1 Created
Attaching to fastapi-app-1
fastapi-app-1 | [nltk_data] Downloading package punkt to /root/nltk_data...
fastapi-app-1 | [nltk_data] Unzipping tokenizers/punkt.zip.
fastapi-app-1 | INFO: Started server process [1]
fastapi-app-1 | INFO: Waiting for application startup.
fastapi-app-1 | INFO: Application startup complete.
fastapi-app-1 | INFO:
                                  Uvicorn running on http://0.0.0.0:8000 (Press CTRL+C to quit)
    ☐ ✓ ≨ fastapi
                                                                        Running (1/1)
```

Running 8000:8000 ☑ 3 minutes ago

Containerisation du frontend avec Docker :

fastapi-app:latest

a) Construction de l'image Docker:

app-1

141b684142ab

```
Dockerfile > ...

# Build stage

FROM node:18.13.0-alpine as build

RUN mkdir -p /app

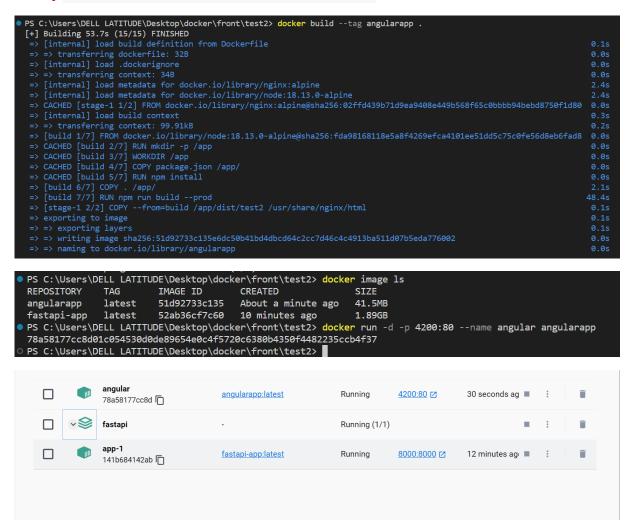
WORKDIR /app

COPY package.json /app/
RUN npm install

COPY . /app/
RUN npm run build --prod

PROM nginx:alpine
COPY --from=build /app/dist/test2 /usr/share/nginx/html
```

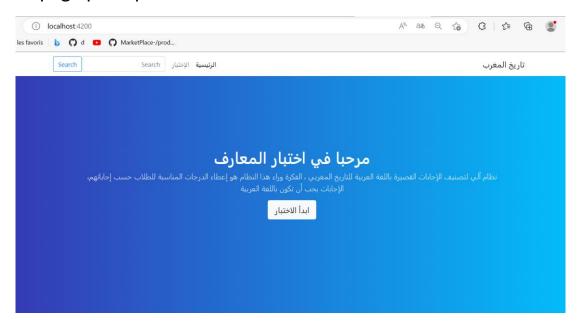
b) Création du conteneur Docker :



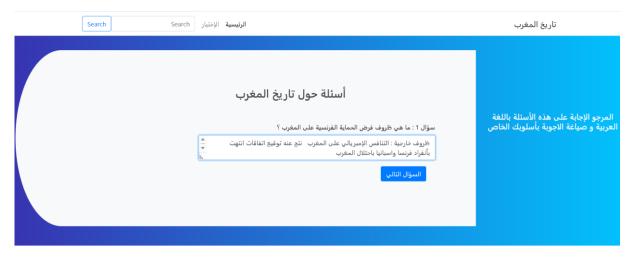
Maintenant on va tester le model après le lancement des deux applications

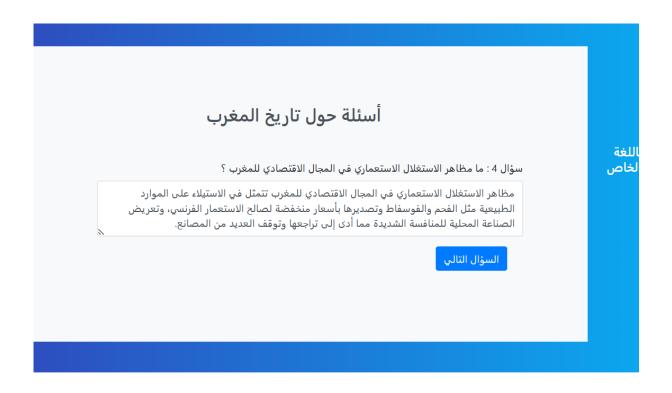
Application

La page principale du site :



La page pour répondre aux 10 questions :

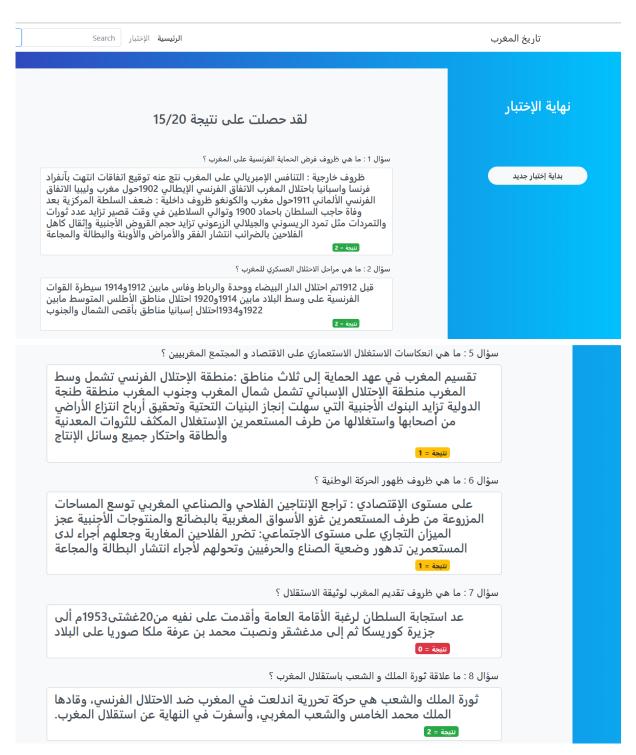






Après avoir répondu à toutes les questions on obtient le résultat total et le résultat pour chaque question :

(Réponses de personne 1)



(Réponses de personne 2)

