

# DETECTEZ LES BAD BUZZ GRÂCE AU DEEP LEARNING

Présentation “Projet 7” chez “OPENCLASSROOM”  
Jaoid KRAIRI  
(Janvier 2021)

# SOMMAIRE



Cahier des charges,



Modèle selon l'approche « API sur étagère »,



Modèle selon l'approche "Modèle sur mesure simple",



Modèles selon l'approche "Modèle sur mesure avancé",



Mise en production du modèle avancé choisi,



Remerciement,

**Cahier des charges:**  
**1/Contexte**

## Rappel du contexte

- ✓ Air Paradis a missionné notre cabinet pour créer un produit IA,
- ✓ Anticiper les bad buzz sur les réseaux sociaux,
- ✓ N'a pas toujours bonne presse sur les réseaux.

## Cahier des charges: 2/ Compte-rendu

- Prédire le sentiment associé à un tweet,
- Pas de données clients chez Air Paradis,
- Utiliser des données Open Source,
- Utilisateur ayant posté, contenu, moment du post et label,
- Préparer un prototype fonctionnel du modèle,
- Préparer un support de présentation explicitant la méthodologie utilisée.

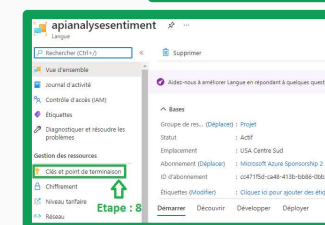
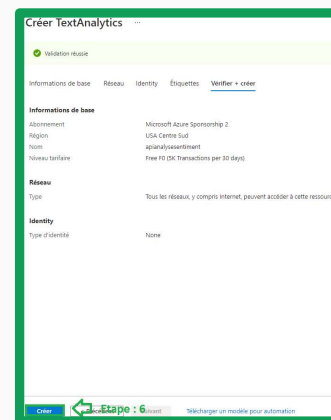
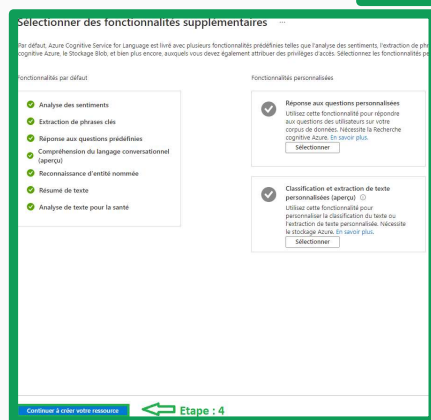
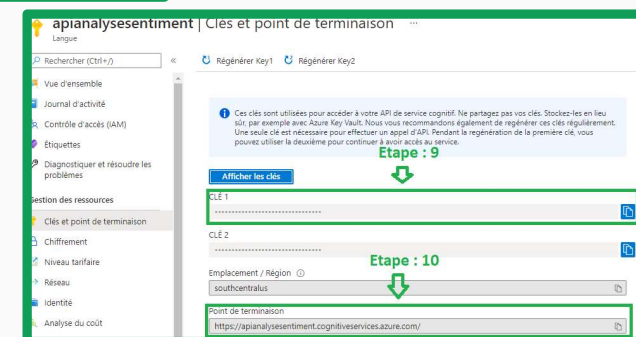
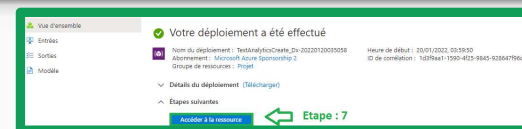
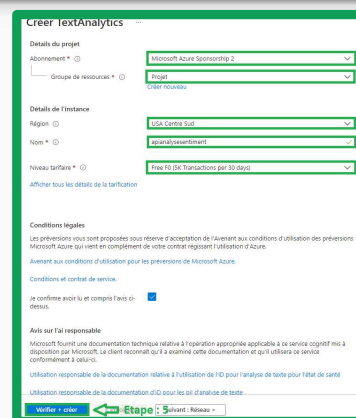
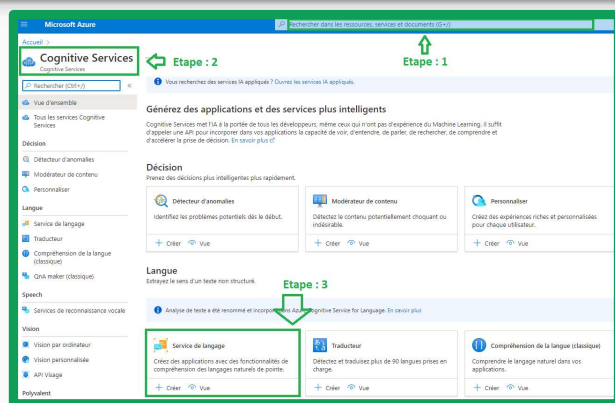
## **Cahier des charges:**

### **3/ Directive de mon manager**

- Marc, mon manager, m'a contacté pour, selon ses mots, "faire d'une pierre deux coups",
- Ce produit pourrait se généraliser à d'autres cas d'usage,
- De tester plusieurs approches,
- L'approche "API sur étagère",
- L'approche "Modèle sur mesure simple",
- L'approche "Modèle sur mesure avancé".

# Modèle selon l'approche « API sur étagère » :

## 1/ Déployer l'API « Analyse de texte » du service cognitif Azure



## Modèle selon l'approche « API sur étagère » :

### 2/ Appeler l'API « Analyse de texte » du service cognitif Azure en local sur un article blog avion

Créer un fichier nommé « P7\_02\_scripts\_api\_analyse\_sentiment »

Entrée [1]: 1 pip install azure-ai-textanalytics==5.1.0

```
import para
key = para.key
endpoint = para.endpoint

from azure.ai.textanalytics import TextAnalyticsClient
from azure.core.credentials import AzureKeyCredential

# M'authentifier à l'aide de ma clé et de mon point de terminaison
def authenticate_client():
    ta_credential = AzureKeyCredential(key)
    text_analytics_client = TextAnalyticsClient(
        endpoint=endpoint,
        credential=ta_credential)
    return text_analytics_client

client = authenticate_client()
```

```
def sentiment_analysis_example(client):

    documents = ["My next plane ticket has just been booked! I fly in March to the Philippines. I"]
    response = client.analyze_sentiment(documents=documents)[0]
    print("Document Sentiment: {}".format(response.sentiment))
    print("Overall scores: positive={0:.2f}; neutral={1:.2f}; negative={2:.2f} \n".format(
        response.confidence_scores.positive,
        response.confidence_scores.neutral,
        response.confidence_scores.negative,
    ))
    for idx, sentence in enumerate(response.sentences):
        print("Sentence: {}".format(sentence.text))
        print("Sentence {} sentiment: {}".format(idx+1, sentence.sentiment))
        print("Sentence score:\nPositive={0:.2f}\nNeutral={1:.2f}\nNegative={2:.2f}\n".format(
            sentence.confidence_scores.positive,
            sentence.confidence_scores.neutral,
            sentence.confidence_scores.negative,
        ))

sentiment_analysis_example(client)
```

```
Document Sentiment: mixed
Overall scores: positive=0.44; neutral=0.09; negative=0.47
```

# Modèle selon l'approche "Modèle sur mesure simple" :

## 1/ Préparation du service Azure

The screenshot shows the Azure Machine Learning Designer interface. On the left is a sidebar with navigation options: New, Home, Author, Notebooks, Automated ML, Designer (highlighted with a red box), Assets, Datasets, Experiments, Pipelines, Models, Endpoints, Manage, Compute, Datastores, and Data labeling. The main area is titled 'docs-ws > Designer' and 'Designer'. It features a 'New pipeline' section with a large blue plus icon in a box (highlighted with a red box) labeled 'Easy-to-use prebuilt modules' and a 'Sample 1: Regression - Automobile Price Prediction...' pipeline. Below this is a 'Pipelines' section with tabs for 'Pipeline drafts' and 'Pipeline runs'. The 'Pipeline drafts' tab shows a table with one entry: 'Sample pipeline' of type 'Training'.

Name	Pipeline type
Sample pipeline	Training

The screenshot shows the Azure Machine Learning Compute interface. The left sidebar lists various components: Nouveau, Accueil, Auteur, Notebooks, ML automatisé, Concepteur, Actifs multimédias, Jeux de données, Expériences, Pipelines, Modèles, Points de terminaison, Gérer, Calcul (selected), Environnements, Magasins de données, Étiquetage des données, and Services liés. The main area is titled 'Calcul' and has tabs for 'Instances de calcul', 'Clusters de calcul' (selected), 'Clusters d'inférence', and 'Calculs attachés'. It features a graphic of server racks and a cloud with a plus sign. Below the graphic is the text: 'Mettre à l'échelle votre cluster de calcul d'une charge de travail mononœud vers une charge de travail multinœud'. Further down, it says: 'Créez un cluster de calcul mono- ou multinœud pour vos charges de travail d'entraînement, d'inférence de lot ou d'apprentissage par renforcement. En savoir plus'. At the bottom, there is a '+ Nouveau' button and a link to 'Voir les tutoriels Azure Machine Learning'.



# Modèle selon l'approche "Modèle sur mesure simple" :

## 2/ Importation du jeu de données de 1000 tweets

Créer un fichier nommé « P7\_02\_scripts\_prepa\_jeu\_mod\_simple »

```
1 print("après:.....")
2 df_1.head(10)
```

après:.....

	label	tweet
0	negative	[cancelled, thinking, bout, watching, movie, d...
1	negative	[andy, avalos, morning, said, would, see, dry, ...
2	negative	[destroy, twitter, similar, tweetdeckno]
3	negative	[sugarland, singing, quotlove, shackquot, like...
4	negative	[yey, good, luck, haha, aww, aryt, see, tomorr...
5	negative	[going, lay, crampy, feeling, yuck, really, ha...
6	negative	[hit]
7	negative	[heard, divorce, family, sort, thing, soooooo, ...
8	negative	[miss, allen]
9	negative	[miss, much]

```
1 # des tweets vides après nettoyage?
2 print(df_1.loc[df_1.tweet==""].shape[0])
```

0

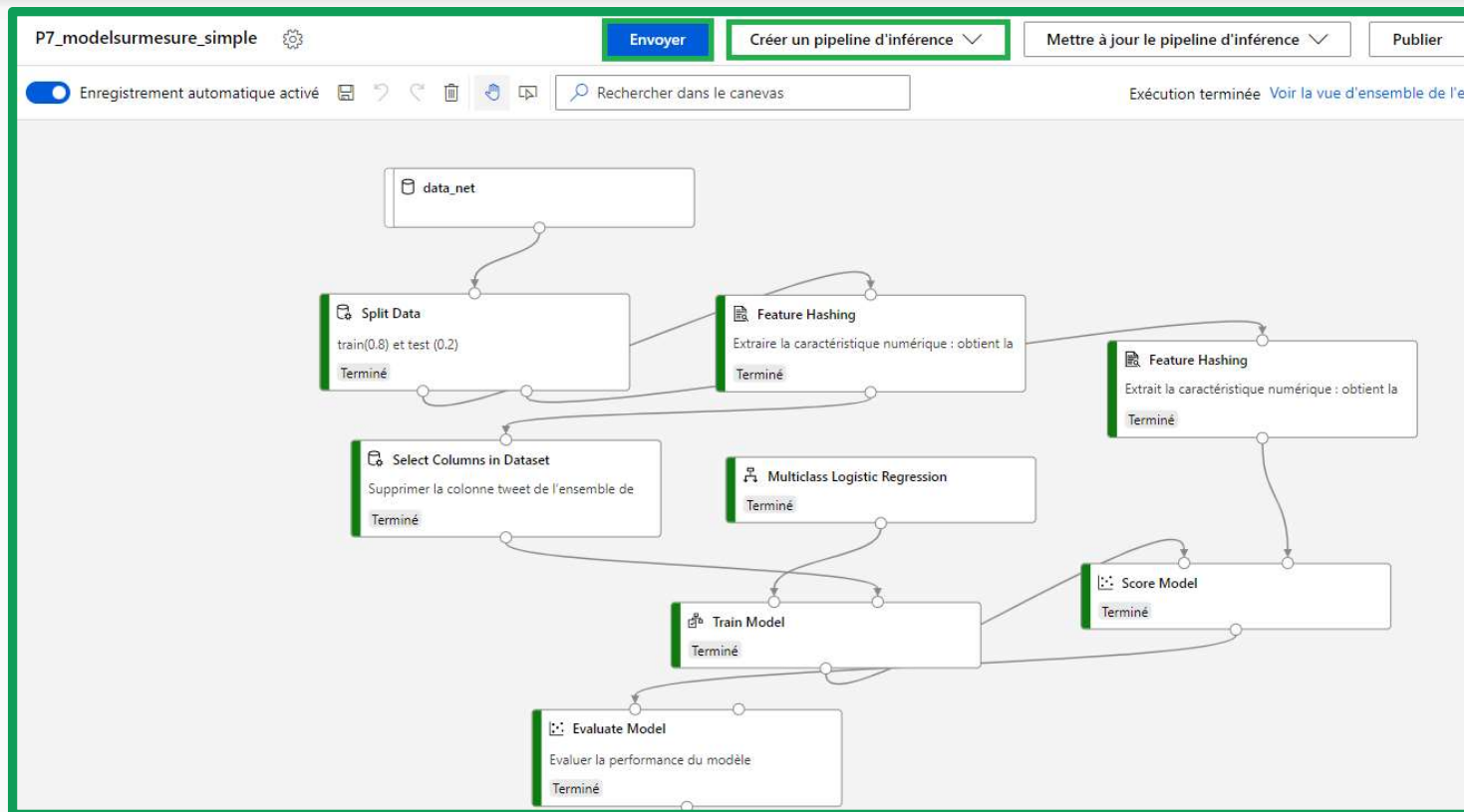
```
1 # retrait des tweets correspondants
2 df_ok = df_1.loc[df_1.tweet != ""]
3 print(df_ok.shape)
```

(1000, 2)

```
1 df_ok.to_csv (r'P7_03_fichiercsv_tweet_red_simple_net_cat_test_1000.csv', index = False, header=True)
```

The screenshot shows the Microsoft Azure Machine Learning Studio interface. The left sidebar contains navigation options: Nouveau, Accueil, Auteurs, Notebooks, ML automatisé, Concepteur, Actifs multimédias, Jeux de données (highlighted), Expériences, Pipelines, Modèles, Points de terminaison, Générer, Calcul, and Environnements. The main area displays details for a dataset named 'data\_net' (Version 1). The 'Attributs' section shows 'Propriétés Tabulaire', 'Créé par Jaoid Krairi', 'Profil Aucun profil généré', 'Fichiers dans le jeu de données 1', and 'Taille totale des fichiers dans le jeu de données 72,99 Kio'. The 'Étiquettes' section shows 'Aucune donnée'. The 'Description' section shows 'jeu de données prétraité et nettoyé'. The 'Sources de données' section shows 'Magasin de données: workspaceblobstore' and 'UI/01-16-2022\_040545\_UTC P7\_03\_fichiercsv\_tweet\_red\_simple\_net\_cat\_test\_1000.csv'.

## Modèle selon l'approche "Modèle sur mesure simple" : 3/ Former le modèle



# Modèle selon l'approche "Modèle sur mesure simple" :


## 4/ Créer un clusters d'inférence

Microsoft Azure Machine Learning Studio

Accueil > Calcul

### Calcul

Instances de calcul   Clusters de calcul   Clusters d'inférence   Calculs attachés



**Déployer votre modèle sur Azure Kubernetes Service pour une inférence à grande échelle**

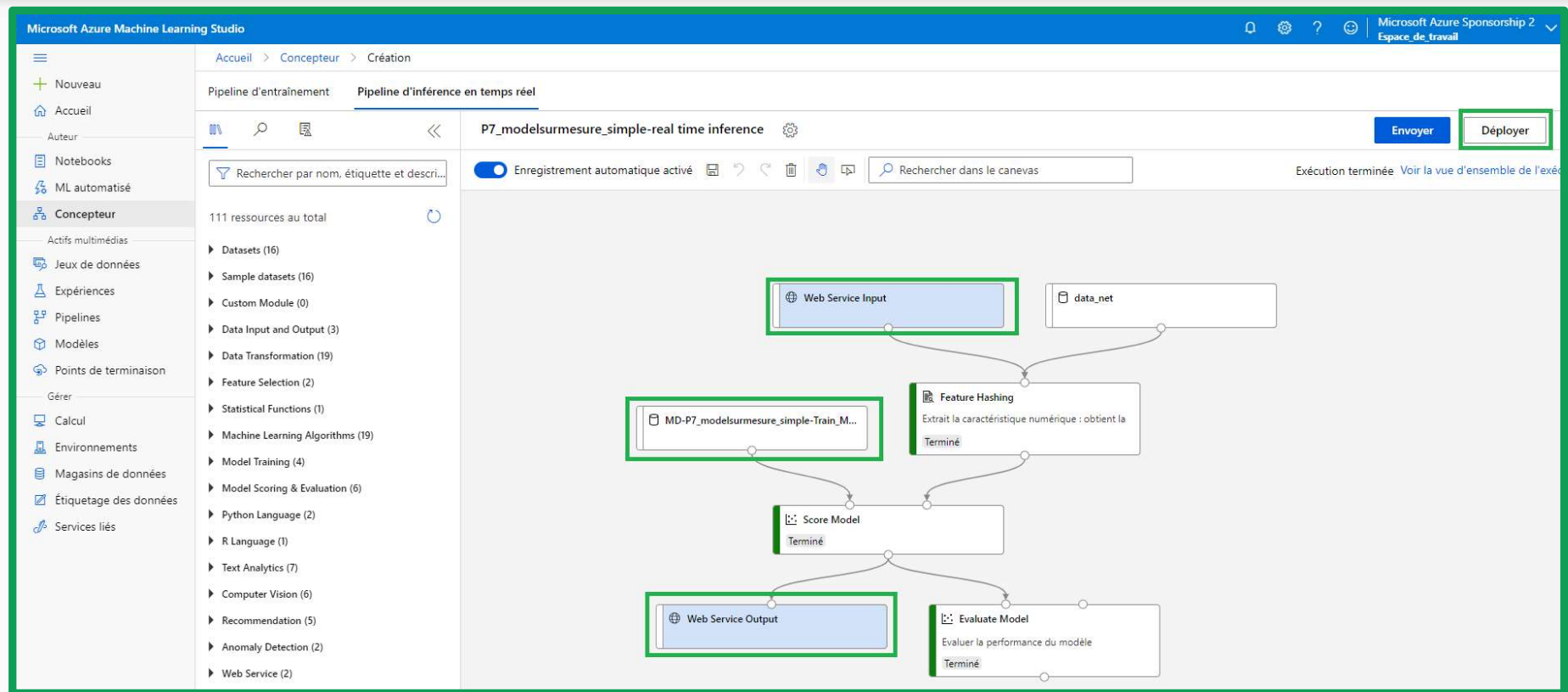
Créez un cluster Azure Kubernetes Service (AKS) ou attachez un cluster AKS existant à votre espace de travail, puis déployez votre modèle comme un point de terminaison REST. [En savoir plus](#)

[+ Nouveau](#)

[Voir les tutoriels Azure Machine Learning](#)

# Modèle selon l'approche "Modèle sur mesure simple" :

## 5/ Création d'une inférence en temps réel



# Modèle selon l'approche "Modèle sur mesure simple" :

## 6/ Le modèle est déployé

The screenshot displays the Microsoft Azure Machine Learning Studio interface. The left sidebar contains navigation options: Nouveau, Accueil, Auteur, Notebooks, ML automatisé, Concepteur, Actifs multimédias, Jeux de données, Expériences, Pipelines, and Modèles (highlighted with a green box). The main area shows the breadcrumb 'Accueil > Modèles > amlstudio-simple:1'. Below this, the 'Points de terminaison' tab is selected and highlighted with a green box. Other tabs include Détails, Versions, Artefacts, Explications (préversion), Justesse (préversion), and Jeux de données. Action buttons for 'Déployer', 'Actualiser', 'Modifier les colonnes', and 'Réinitialiser la vue' are visible. A search bar labeled 'Rechercher' is present. The table below shows the endpoint details:

Nom	Description	Créé le	Créé par	Mis à jour le ↓	Type de calcul	Cible de calcul	Étiquettes
simple	déterminer le sentiment d'u...	16 janv. 2022 06:03	Jaoid Krairi	16 janv. 2022 06:03	Kubernetes	dd	CreatedByAMLStudio: true ...

# Modèle selon l'approche "Modèle sur mesure simple" :

## 7/ Tester le modèle sur un article de blog

Accueil > Points de terminaison > simple

### simple

Détails **Test** Consommer Journaux de déploiement

Entrer des données pour tester le point de terminaison en temps réel

Sélectionner le type d'éditeur

☒ Éditeur de formulaire ☐ Éditeur JSON

WebServiceInput0

**Test** Lancement du test

Résultat de test

Le résultat de test est trop long pour être affiché. Longueur : 57810

analysés **bruts** Cliquez pour copier

J'ai copié le résultat que j'ai enregistré dans un bloc note au format texte

label

negative

Je détermine arbitrairement un label négatif

tweet

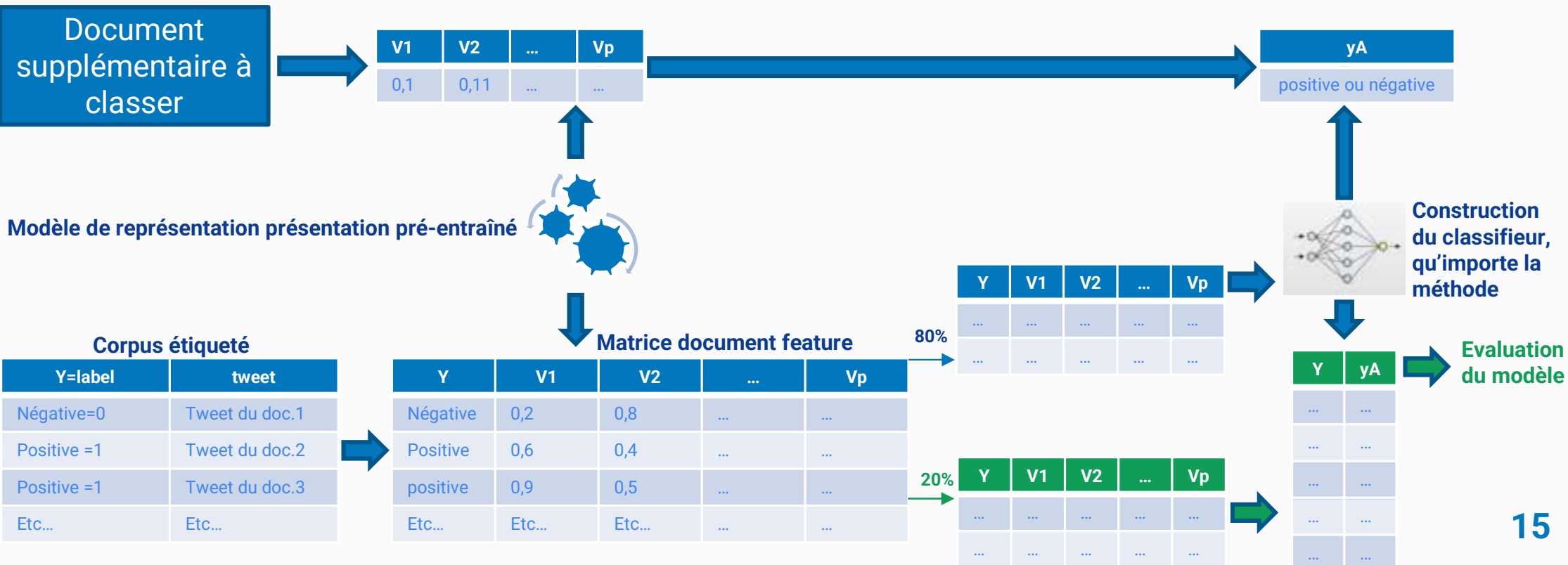
My next plane ticket has just been booked! I fly in March to ...

J'intègre mon article de blog d'avion de plus de 800 mots environ

```
"Scored Probabilities negative": 0.46346136371179714,  
"Scored Probabilities positive": 0.5365386362882029,  
"Scored Labels": "positive"
```

# Modèles selon l'approche "Modèle sur mesure avancé" :

## 1/ Description générale de la méthode



# Modèles selon l'approche "Modèle sur mesure avancé" :

## 2/ Exemple hyperparamètres du modèle keras embedding optimisé

### Modélisation prédictive (1) - régularisation KERAS simple(optimisé)

```
1 # créer un modèle
2 from tensorflow.keras.models import Sequential
3 from tensorflow.keras.layers import Dense, Flatten, Embedding, MaxPooling1D, Dropout
4 # perception multicouche
5 #==> output_dim précise la taille de l'espace de représentation dans laquelle seront projetés les termes
6 #==> input_dim = dico_size + 1 à cause de l'index des termes commence à la colonne 1
7 # (la colonne 0 existe mais n'est pas associée à un terme)
8 model_KSO = Sequential()
9 model_KSO.add(Embedding(input_dim = dico_size + 1, output_dim = 10, input_length = max_length + marge_length))
10 # 2 opérations de régularisation
11 model_KSO.add(MaxPooling1D(pool_size=5, strides=2))
12 model_KSO.add(Dropout(0.7))
13 # reprise du réseau habituel
14 model_KSO.add(Flatten())
15 model_KSO.add(Dense(units=1, activation="sigmoid"))
16
17 # structure du réseau
18 print(model_KSO.summary())
```

Model: "sequential"

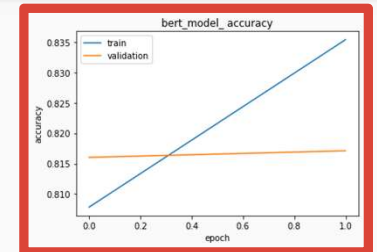
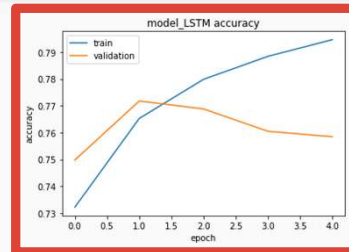
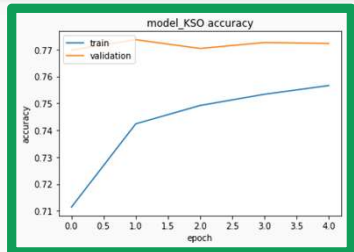
Layer (type)	Output Shape	Param #
embedding (Embedding)	(None, 38, 10)	3193600
max_pooling1d (MaxPooling1D)	(None, 17, 10)	0
dropout (Dropout)	(None, 17, 10)	0
flatten (Flatten)	(None, 170)	0
dense (Dense)	(None, 1)	171

=====  
Total params: 3,193,771  
Trainable params: 3,193,771  
Non-trainable params: 0



# Modèles selon l'approche "Modèle sur mesure avancé" :

## 3/ Comparer la performance des 5 modèles



Modèle	Taille du jeu de données(%)	Perte(%)	Performance(%)	Durée formation modèle (heure)	Durée évaluation modèle (seconde)
model_KSO (Keras Embedding)	100%	49,58%	77,03%	0h14min et 8 secondes	43 secondes
model_LSTM (Keras Embedding avec LSTM)	100%	50,41%	76,05%	0h23min et 6 secondes	68 secondes
bert_model	100%	41,34%	81,66%	05h04min et 51 secondes	935 secondes
clf ( Word2Vec SVM avec noyau RBF par défaut)	2,5%	Non indiqué	70%	02h	Non indiqué
clf ( glove SVM avec noyau RBF par défaut)	2,5%	Non indiqué	74%	02h	Non indiqué

# Mise en production du modèle avancé choisi :

## 1/Cartographie

1

Connection à mon espace de travail

2

Créer une expérience

3

Joindre à l'instance de calcul

4

Charger les données du datastore

5

Créer un répertoire

6

Script de formation

7

Créer un script de configuration d'exécution

8

Créer un ScriptRunConfig

9

Soumettre le travail

10

Enregistrer le modèle

11

Fonction d'appel

12

Créer un fichier de configuration

13

Déploiement dans l'ACI

# REMERCIEMENT

Merci de m'avoir écouté

## REPONDRE AUX QUESTIONS