#CLASSIFICATION : ENTITES NOMMEES-TRUE_FALSE_OTHER_MIXTURE :

Membres: Hadjoudja Bachir (21811363), Zeggar Rym (21909615), Bendahmane Rania (21811387), Labiad Youcef (21710780).

```python
#les imports utilisés
import sys
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
import re
import string
import time
import warnings
import nltk
from nltk.corpus import stopwords
from nltk.stem import WordNetLemmatizer
from nltk.stem.porter import PorterStemmer
from nltk import word_tokenize
import spacy
from spacy import displacy
from sklearn.datasets import fetch_20newsgroups
from sklearn.preprocessing import LabelEncoder
from sklearn.feature_extraction.text import CountVectorizer,
TfidfVectorizer
from sklearn.model_selection import train_test_split, GridSearchCV,
KFold, cross_val_score
from sklearn.pipeline import Pipeline
from sklearn.metrics import accuracy_score, confusion_matrix,
classification_report, precision_recall_fscore_support as score
from sklearn.metrics import ConfusionMatrixDisplay
from sklearn.naive_bayes import GaussianNB, MultinomialNB
from sklearn.linear_model import LogisticRegression
from sklearn.tree import DecisionTreeClassifier
from sklearn.neighbors import KNeighborsClassifier
from sklearn.svm import SVC
from sklearn.ensemble import RandomForestClassifier
from torch.utils.data import Dataset, DataLoader, random_split
from torch import Tensor
from torch.nn import Linear, ReLU, Sigmoid, Module, BCELoss
from torch.optim import SGD
from torch.nn.init import kaiming_uniform_, xavier_uniform_
from tabulate import tabulate
from google.colab import drive
# Sickit learn met régulièrement à jour des versions et indique des
futurs warnings.
# Ces deux lignes permettent de ne pas les afficher.
warnings.filterwarnings("ignore", category=FutureWarning)
```

```python
# Téléchargez les données nécessaires pour certaines fonctions de
traitement du texte en utilisant nltk
nltk.download('stopwords')
nltk.download('punkt')
nltk.download('wordnet')
```

```
[nltk_data] Downloading package stopwords to /root/nltk_data...
[nltk_data]   Unzipping corpora/stopwords.zip.
[nltk_data] Downloading package punkt to /root/nltk_data...
[nltk_data]   Unzipping tokenizers/punkt.zip.
[nltk_data] Downloading package wordnet to /root/nltk_data...

True
```

autorisation

```python
drive.mount('/content/gdrive/')
```

```
Mounted at /content/gdrive/
```

```python
my_local_drive='/content/gdrive/My Drive'
# Ajout du path pour les librairies, fonctions et données
sys.path.append(my_local_drive)
# Se positionner sur le répertoire associé
%cd $my_local_drive
%ls

%pwd
```

```
/content/gdrive/My Drive
 ⓖⓇⓖⓇ/
'ⓟⓛ '/
 0341_001.pdf
 0341_001.pdf.gdoc
 141APPLE/
 20220513_122210.jpg
'ADHE-2020-a0ad-1 (1).pdf'
 ADHE-2020-a0ad-1.pdf
 AGL/
 ARTICLE.gdoc
'attestation d employeur 001.jpg'
'ATTESTATION-DE-PISE-EN-CHARGE ++.docx'
 attestation_hebergement-rania.pdf
 attestationnn_d_hebergement_etudiant.pdf
 Avis_d_impot_2020_sur_les_revenus_2019.pdf
 Banque/
'BENDAHMANE RANIA.gdoc'
 Bershka_40117332158.gdoc
 Bershka_40117332158.pdf
'BIGDATA OFFICIEL.gdoc'
 Bruxelles/
'CamScanner 04-15-2022 12.41.pdf'
```

```
'CamScanner 09-22-2020 12.22.30_1.pdf'
'cas raphia.pdf'
'CCF17112019_00000 (1).jpg'
 CCF17112019_00000.jpg
'CCI17112019_00000 (1).bmp'
 CCI17112019_00000.bmp
'CCI17112019_00001 (1).bmp'
 CCI17112019_00001.bmp
'CCI17112019_00002 (1).bmp'
 CCI17112019_00002.bmp
 CCI17112019_00003.bmp
 CCI17112019_00004.bmp
'CCI17112019_00005 (1).bmp'
 CCI17112019_00005.bmp
'Chill_Tea Projet 2020-2021'/
 CI_REO_dossier_inscriptions.pdf
'Colab Notebooks'/
 COMBI/
'Copie de FakeNEWS.ipynb'
'correction arithmetique.zip'
'Courrier MAE.gdoc'
'Courrier MAE.pdf'
 CR_ALGEBRE_NOTE.pdf
 Creation_groupeA.zip
'C&T - Modèle relationnel BDD .gdoc'
 CV.gdoc
'Cv stage pharmacie.pptx'
'CY20029073_PACKEXCEPTION (1).pdf'
 CY20029073_PACKEXCEPTION.gdoc
 CY20029073_PACKEXCEPTION.pdf
 Cyprus/
'Demandes de CP - congés payés (1).doc'
'Demandes de CP - congés payés (1).docx'
'Demandes de CP - congés payés (1).gdoc'
'Demandes de CP - congés payés (2).gdoc'
'Demandes de CP - congés payés.doc'
'Demandes de CP - congés payés.docx'
'Demandes de CP - congés payés.gdoc'
'Demandes de CP - congés payés Rania.doc'
'Demandes de CP - congés payés Rania.gdoc'
'Document sans titre (10).gdoc'
'Document sans titre (11).gdoc'
'Document sans titre (12).gdoc'
'Document sans titre (13).gdoc'
'Document sans titre (14).gdoc'
'Document sans titre (15).gdoc'
'Document sans titre (16).gdoc'
'Document sans titre (17).gdoc'
'Document sans titre (18).gdoc'
'Document sans titre (19).gdoc'
```

```
'Document sans titre (1).gdoc'
'Document sans titre (20).gdoc'
'Document sans titre (21).gdoc'
'Document sans titre (22).gdoc'
'Document sans titre (23).gdoc'
'Document sans titre (24).gdoc'
'Document sans titre (25).gdoc'
'Document sans titre (26).gdoc'
'Document sans titre (27).gdoc'
'Document sans titre (28).gdoc'
'Document sans titre (29).gdoc'
'Document sans titre (2).gdoc'
'Document sans titre (30).gdoc'
'Document sans titre (31).gdoc'
'Document sans titre (32).gdoc'
'Document sans titre (33).gdoc'
'Document sans titre (34).gdoc'
'Document sans titre (35).gdoc'
'Document sans titre (36).gdoc'
'Document sans titre (37).gdoc'
'Document sans titre (38).gdoc'
'Document sans titre (39).gdoc'
'Document sans titre (3).gdoc'
'Document sans titre (40).gdoc'
'Document sans titre (41).gdoc'
'Document sans titre (42).gdoc'
'Document sans titre (43).gdoc'
'Document sans titre (44).gdoc'
'Document sans titre (45).gdoc'
'Document sans titre (46).gdoc'
'Document sans titre (47).gdoc'
'Document sans titre (48).gdoc'
'Document sans titre (49).gdoc'
'Document sans titre (4).gdoc'
'Document sans titre (5).gdoc'
'Document sans titre (6).gdoc'
'Document sans titre (7).gdoc'
'Document sans titre (8).gdoc'
'Document sans titre (9).gdoc'
'Document sans titre.gdoc'
'Document sans titre - Google Docs.pdf'
'Document sans titre.pdf'
'DROP DATABASE IF EXISTS SportPlannerBD;.gdoc'
 DSC_0040.jpg
 DSC_0048.jpg
'DSC_0089 (1).jpg'
 DSC_0089.jpg
'DSC_0090 (1).jpg'
 DSC_0090.jpg
'DSC_0092 (1).jpg'
```

```
 DSC_0092.jpg
 DSC_0093.jpg
 DSE.pdf
'E11ARY5 G1G3 L1 S1 EV1.gdoc'
'E11ARY5 G1G3 L1 S1 EV1.odt'
'E14ARY5 arabophones et non arabophones 1ère session (2).docx'
'E14ARY5 arabophones et non arabophones 1ère session (2).gdoc'
'enercoop 001.jpg'
'enercoop 2 001.jpg'
'english presenation .gdoc'
'ENGLISH PRESENTATION ,.gdoc'
'EVALUATION MANAGEMENT L2 FDS partie 2.docx'
'EVALUATION MANAGEMENT L2 FDS partie 2.gdoc'
'Exam HLEE202 session1 250520 (1).gdoc'
'Exam HLEE202 session1 250520.doc'
'Exam HLEE202 session1 250520.gdoc'
 exo7tp/
"facture d'éléctricité.pdf"
'Feuille de présence auxiliaire de v (1).xlsx'
'Feuille de présence auxiliaire de v (2).xlsx'
'Feuille de présence auxiliaire de v.xlsx'
'Feuille de presence intervenant (1).gdoc'
'Feuille de presence intervenant (1).pdf'
'Feuille de presence intervenant (2).gdoc'
'Feuille de presence intervenant (2).pdf'
'Feuille de presence intervenant (3).pdf'
'Feuille de presence intervenant (4).pdf'
'Feuille de presence intervenant.gdoc'
'Feuille de presence intervenant.pdf'
'FICHE DE RENSEIGNEMENTS.gdoc'
'FICHE DE RENSEIGNEMENTS.pdf'
 fiche_famille.pdf
'FICHES DE PAIE.zip'
 FP.gdoc
'Fwd Chemsseddine Hadiby.zip'
'Grille_auto_eval_v2 (1).gsheet'
'Grille_auto_eval_v2 (1).ods'
'Hadjoudja Bachir.docx'
 HLIN406ccMars2021.gdoc
'iae tableau.gdoc'
 IA_FRM_Declaration_honneur_justificatif_domicile_V1-1-1.docx
 ia.gdoc
'ia Projet.gdoc'
'identite 1 001.jpg'
'identite2 001.jpg'
 ilovepdf_merged_compressed.pdf
 ilovepdf_merged.pdf
'image3 (1).jpeg'
 IMG_1276.JPG
'impots 1 001.jpg'
```

'impots 2 001.jpg'
'imprime_da_general (1).gdoc'
'imprime_da_general (2).gdoc'
'imprime_da_general (3).gdoc'
 imprime_da_general.gdoc
 imprime_da_general.pdf
'Imprimé DA (informatique).gdoc'
'Imprimé DA (informatique).pdf'
 informations.pdf
 instructions.pdf
 jena/
'L2 S6'/
'L3 Info D.xls'
 Lettre-de-motivation-assistante-sociale-Crous.gdoc
 Lettre-de-motivation-assistante-sociale-Crous.pdf
 Lettre.gdoc
'Liste 6 des inscrits UES MAPS et BCM.pdf'
'Luxembourg '/
 ML/
'modèle relationnel tournoi(1).gdoc'
'modèle relationnel tournoi(1).pdf'
'MP Attestation de loyer recto_Attestation résidence verso (homologué
11_2013)_MP Attestation de loyer Recto_Attestation résiden.pdf'
 newsTrain2.csv
 newsTrain_-_newsTrain.csv
'Non imposition (1).gdoc'
'Non imposition.gdoc'
'Non imposition.pdf'
'ok (1).jpg'
 ok.jpg
'ou la mauvaise perception de soi conduit à des comportements
inadaptés qui renforcent la mauvaise image de soi , inversement une
bonn.gdoc'
'paye 001.jpg'
'paye 2 001.jpg'
'paye juin 001.jpg'
'paye juin2 001.jpg'
'paye mai 001.jpg'
'paye mai 2 001.jpg'
'Planning mensuel intervenant (1).pdf'
'Planning mensuel intervenant.pdf'
'Poland Rania'/
'POO Cours 7.webm'
'POO TP_TD 6(nebut).webm'
'Présentation tournoi.gslides'
 PROJET_GROUPEC3/
'Quittance 2018-05-01.pdf'
 Raniaparis/
'Rapport Final PPE.docx'
'Rapport Final PPE.gdoc'

```
'RIB 001.jpg'
 Rome/
'Rome (1)'/
 S4-20210504T124747Z-001.zip
 S5-20210504T123656Z-001.zip
 S6/
'S6 (1)'/
 S6-20210504T123706Z-001.zip
'\section{Complétion de l'\''ontologie}.gdoc'
'*.tar.gz'
 TP/
 TP2EXO5/
 TP4et5multitaches/
 TPexo4/
 tp_obwan/
'TP TP PROG APPLICATIVE'/
 traitement.gdoc
باريز/
```

{"type":"string"}

La fonction qui sera utilisée pour les prétraitements: MyCleanText

- Mettre le texte en minuscule
- Se débarasser des stopwords
- Se débarasser des nombres
- Stemmatisation
- Lemmatisation ..

La fonction MyshowAllScores prend le y_test et le y_predict, affiche l'accuracy et le classification report avec la matrice de confusion.

```
#................................................Fonction
MyCleanText ..................................................
....................
# mettre en minuscule
#enlever les stopwords
#se debarasser des nombres
#stemmatisation
#lemmatisation
#...........................................................
.............................................................
...........


nltk.download('wordnet')
nltk.download('stopwords')
nltk.download('punkt')
#liste des stopwords en anglais
stop_words = set(stopwords.words('english'))
```

```python
def MyCleanText(X,
                lowercase=False, #mettre en minuscule
                removestopwords=False, #supprimer les stopwords
                removedigit=False, #supprimer les nombres
                getstemmer=False, #conserver la racine des termes
                getlemmatisation=False #lemmatisation des termes
                ):
    #conversion du texte d'entrée en chaîne de caractères
    sentence=str(X)
    #suppression des caractères spéciaux
    sentence = re.sub(r'[^\w\s]',' ', sentence)
    # suppression de tous les caractères uniques
    sentence = re.sub(r'\s+[a-zA-Z]\s+', ' ', sentence)
    # substitution des espaces multiples par un seul espace
    sentence = re.sub(r'\s+', ' ', sentence, flags=re.I)

    # decoupage en mots
    tokens = word_tokenize(sentence)
    if lowercase:
        tokens = [token.lower() for token in tokens]

    # suppression ponctuation
    table = str.maketrans('', '', string.punctuation)
    words = [token.translate(table) for token in tokens]

    # suppression des tokens non alphabetique ou numerique
    words = [word for word in words if word.isalnum()]

    # suppression des tokens numerique
    if removedigit:
        words = [word for word in words if not word.isdigit()]

    # suppression des stopwords
    if removestopwords:
        words = [word for word in words if not word in stop_words]

    # lemmatisation
    if getlemmatisation:
        lemmatizer=WordNetLemmatizer()
        words = [lemmatizer.lemmatize(word)for word in words]


    # racinisation
    if getstemmer:
        ps = PorterStemmer()
        words=[ps.stem(word) for word in words]

    sentence= ' '.join(words)
```

```python
        return sentence

def MyshowAllScores(y_test,y_pred):
    classes= np.unique(y_test)
    print("Accuracy : %0.3f"%(accuracy_score(y_test,y_pred)))
    print("Classification Report")
    print(classification_report(y_test,y_pred,digits=5))
    cnf_matrix = confusion_matrix(y_test,y_pred)
    disp=ConfusionMatrixDisplay(cnf_matrix,display_labels=classes)
    disp.plot()
```

```
[nltk_data] Downloading package wordnet to /root/nltk_data...
[nltk_data] Downloading package stopwords to /root/nltk_data...
[nltk_data]   Unzipping corpora/stopwords.zip.
[nltk_data] Downloading package punkt to /root/nltk_data...
[nltk_data]   Unzipping tokenizers/punkt.zip.
```

La classe TextNormalizer qui contiendra la fonction MyCleanText.

Fit_transform de mon corpus propre.

```python
#......................................................Etape 1 :
prétraitement du
texte ...................................................................
......................................
#......................................................Class
TextNormalizer  ...................................................................
.........................
#fit_transform de mon corpus propre
#...................................................................
...................................................................
...........

from sklearn.base import BaseEstimator, TransformerMixin

class TextNormalizer(BaseEstimator, TransformerMixin):
    def __init__(self,
                removestopwords=False, # suppression des stopwords
                lowercase=False,# passage en minuscule
                removedigit=False, # supprimer les nombres
                getstemmer=False,# racinisation des termes
                getlemmatisation=False # lemmatisation des termes
                ):

        self.lowercase=lowercase
        self.getstemmer=getstemmer
        self.removestopwords=removestopwords
        self.getlemmatisation=getlemmatisation
        self.removedigit=removedigit
```

```python
    def transform(self, X, **transform_params):
        # Nettoyage du texte
        X=X.copy() # pour conserver le fichier d'origine
        return [MyCleanText(text,lowercase=self.lowercase,
                            getstemmer=self.getstemmer,
                            removestopwords=self.removestopwords,
                            getlemmatisation=self.getlemmatisation,
                            removedigit=self.removedigit) for text in
X]

    def fit(self, X, y=None, **fit_params):
        return self

    def fit_transform(self, X, y=None, **fit_params):
        return self.fit(X).transform(X)

    def get_params(self, deep=True):
        return {
            'lowercase':self.lowercase,
            'getstemmer':self.getstemmer,
            'removestopwords':self.removestopwords,
            'getlemmatisation':self.getlemmatisation,
            'removedigit':self.removedigit
        }

    def set_params (self, **parameters):
        for parameter, value in parameters.items():
            setattr(self,parameter,value)
        return self
```

1ere classification : True/False VS Mixture/OTHER

Etape 1 : Préparer les données

- Charger et préparer les données à partir des 2 fichiers csv
- Affichages pour tester si cela a bien été fait

```python
dftrain1 = pd.read_csv("/content/gdrive/MyDrive/newsTrain2.csv",
names=['id','text','title','rating'], header=0,sep=',',
encoding='utf8')
dftrain1.reset_index(drop = True, inplace = True)

dftrain2 = pd.read_csv("/content/gdrive/MyDrive/newsTrain_-
_newsTrain.csv", names=['id','text','title','rating'],
header=0,sep=',', encoding='utf8')
dftrain2.reset_index(drop = True, inplace = True)


# concaténer les deux dataframes en ajoutant les lignes du deuxième à
# la fin du premier
dftrainbase = pd.concat([dftrain1, dftrain2], ignore_index=True)
```

```python
dftrain=dftrainbase

#On crée une colonne regroupe qui va mettre dans les lignes là où a
true ou bien false la valeur TRUE/FALSE et OTHER ça laisse
#dftrain['truefalse'] = dftrain['rating'].apply(lambda x: 'TRUE/FALSE'
if x in ['TRUE', 'FALSE'] else '')
#On crée une colonne regroupe qui va mettre dans les lignes là où a
true ou bien false la valeur TRUE/FALSE et OTHER ça laisse
#dftrain['mixtureother'] = dftrain['rating'].apply(lambda
x:'MIXTURE/OTHER' if x in ['MIXTURE', 'OTHER'] else '')

dftrain['regrouped'] = dftrain['rating'].apply(lambda x:'TRUE/FALSE'
if x in ['TRUE', 'FALSE'] else 'MIXTURE/OTHER')




print("Echantillon de mon dataset \n")
print(dftrain.sample(n=10))
print("\n")
print("Quelques informations importantes \n")
dftrain.info()
print("\n")
X_text=dftrain["text"]
X_title=dftrain["title"]


print("le texte est")
display(X_text)
print("\n")
print("le titre est")
display(X_title)
print("\n")
y=dftrain.iloc[0:,-1]
print("voici la dernière case de rating")
display(y)
print("\n")
print("la taille de Xt est",X_text.shape)
print("\n")
print(" y EST " ,y)
print("\n")
y = y.str.lower()
print("Les valeurs de true et false sont:\n", y.value_counts())
```

Echantillon de mon dataset

	id	text
375	6a72f5c5	James Humphrey Terwilliger Woods, conservative...
2209	721fde78	FOR IMMEDIATE RELEASE: Contact: Jennie Kennedy...
2127	6378a0ff	A coalition of civil society groups has descri...
1787	8.00E+25	A coalition of civil society groups has descri...

```
1746  b4b6057f  SHARE Madison - Gov. Scott Walker's proposal t...
268   a85f1aac  US Under Military Law, Congress Arrested, War ...
608   53d111d0  The primary school population in England has c...
841   f1ad8b82  Democrats have said they only intend to restor...
736   43174fe2  What is occurring on Wall Street right now is ...
2212  0c5a27ba  top cancer researchers harvard epa agree fluor...

                                                      title    rating  \
375   James Woods, Actor, Patriot, and Trump Support...    FALSE
2209  Only 3% of Trump Voters Believe Joe Biden Is L...  mixture
2127                                                NaN     TRUE
1787  Tesco food waste rose to equivalent of 119m me...     TRUE
1746          Wages back above pre-economic crisis levels    FALSE
268   US Under Military Law, Congress Arrested, War ...    FALSE
608                                                 NaN  mixture
841                        Texans Breathe a Sigh of Relief  mixture
736   Recent $2M Jackpot Winner Leroy Fick Still Use...  mixture
2212  A Report on Updated Overtime and Food Expenses...  mixture

          regrouped
375      TRUE/FALSE
2209  MIXTURE/OTHER
2127     TRUE/FALSE
1787     TRUE/FALSE
1746     TRUE/FALSE
268      TRUE/FALSE
608   MIXTURE/OTHER
841   MIXTURE/OTHER
736   MIXTURE/OTHER
2212  MIXTURE/OTHER


Quelques informations importantes

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2528 entries, 0 to 2527
Data columns (total 5 columns):
 #   Column     Non-Null Count  Dtype
---  ------     --------------  -----
 0   id         2528 non-null   object
 1   text       2528 non-null   object
 2   title      2482 non-null   object
 3   rating     2528 non-null   object
 4   regrouped  2528 non-null   object
dtypes: object(5)
memory usage: 98.9+ KB


le texte est
```

```
0       Distracted driving causes more deaths in Canad...
1       Missouri politicians have made statements afte...
2       Home Alone 2: Lost in New York is full of viol...
3       But things took a turn for the worse when riot...
4       It's no secret that Epstein and Schiff share a...
                            ...
2523    More than four million calls to the taxman are...
2524    More under-18s are being taken to court for se...
2525    The Government's much vaunted Help to Buy Isa ...
2526    The late Robin Williams once called cocaine "G...
2527    The late Robin Williams once called cocaine "G...
Name: text, Length: 2528, dtype: object
```

le titre est

```
0       You Can Be Fined $1,500 If Your Passenger Is U...
1            Missouri lawmakers condemn Las Vegas shooting
2       CBC Cuts Donald Trump's 'Home Alone 2' Cameo O...
3       Obama's Daughters Caught on Camera Burning US ...
4       Leaked Visitor Logs Reveal Schiff's 78 Visits ...
                            ...
2523    Taxman fails to answer four million calls a ye...
2524    Police catch 11-year-olds being used to sell d...
2525    Help to Buy Isa scandal: 500,000 first-time bu...
2526              A coke-snorting generation of hypocrites
2527              A coke-snorting generation of hypocrites
Name: title, Length: 2528, dtype: object
```

voici la dernière case de rating

```
0           TRUE/FALSE
1         MIXTURE/OTHER
2         MIXTURE/OTHER
3           TRUE/FALSE
4           TRUE/FALSE
              ...
2523        TRUE/FALSE
2524        TRUE/FALSE
2525        TRUE/FALSE
2526        TRUE/FALSE
2527        TRUE/FALSE
Name: regrouped, Length: 2528, dtype: object
```

la taille de Xt est (2528,)

```
 y EST  0           TRUE/FALSE
1        MIXTURE/OTHER
2        MIXTURE/OTHER
3          TRUE/FALSE
4          TRUE/FALSE
              ...
2523       TRUE/FALSE
2524       TRUE/FALSE
2525       TRUE/FALSE
2526       TRUE/FALSE
2527       TRUE/FALSE
Name: regrouped, Length: 2528, dtype: object
```

```
Les valeurs de true et false sont:
 true/false       1578
mixture/other      950
Name: regrouped, dtype: int64
```

Le jeu de données étant déséquilibré, on a pensé à appliquer le downsampling pour équilibrer nos données. on séléctionne des lignes aléatoirement de TRUE/FALSE de telle sorte que le nombre de lignes de TRUE/FALSE soit = au nbr de lignes de MIXTURE/OTHER. et on mélange le DataFrame

```python
#On applique du sous-échantillonnage (downsampling) : car on a plus de
FALSE (578) que des TRUE (211)
# Séparer les classes en deux dataframes
df_false_true = dftrain[dftrain['regrouped'] == 'TRUE/FALSE']
df_other_mixture = dftrain [dftrain['regrouped'] == 'MIXTURE/OTHER']

# Sous-échantillonner la classe majoritaire (FALSE) pour obtenir un
nombre égal d'échantillons pour chaque classe
df_subsampled = df_false_true.sample(n=len(df_other_mixture),
random_state=42)

# Concaténer les deux dataframes
dftrain = pd.concat([df_subsampled, df_other_mixture])

# Mélanger aléatoirement les données
dftrain = dftrain.sample(frac=1, random_state=42)

X_text=dftrain.iloc[0:,1:2]
X_title=dftrain.iloc[0:,2:3]

print("le texte est")
display(X_text)
print("le titre est")
display(X_title)
```

```python
y=dftrain.iloc[0:,-1]
print("le y est")
display(y)
print("la taille de X_text est",X_text.shape)
print("la taille de y_train est " ,y.shape)
print("les valeurs de TRUE et FALSE maintenant sont
" ,y.value_counts())
```

le texte est

```
                                                    text
847    Also on the naughty list were Christmas carols...
447    The number of people applying for UK universit...
2243   Unvaccinated Children Pose No Risk To Anyone, ...
1202   Australia's Great Barrier Reef has worst coral...
1059   One in ten patients admitted to hospital are s...
...                                                  ...
637    As Democrats basked in the glory of their take...
1022   HON. STEFAN I. MYCHAJLIW ERIE COUNTY COMPTROLL...
1515   We've reported on numerous events identified i...
1432   The 2021 Act on Climate, legislation sponsored...
628    Antibody tests support what's been obvious: Co...

[1900 rows x 1 columns]
```

le titre est

```
                                                   title
847    Principal banned candy canes because 'J' shape...
447    On the Capitol: Part-time pay for part-time Le...
2243   One of the strengths—and weaknesses—of a Weste...
1202   Australia's Great Barrier Reef has worst coral...
1059   Hospitals deluged by 5,000 diabetics a DAY: On...
...                                                  ...
637                  UK inflation soars to two-year high
1022                                                  NaN
1515   BREAKING EXCLUSIVE: System 'Glitch' Also Uncov...
1432             State of Rhode Island General Assembly
628          President Buhari Never Said he will Resign

[1900 rows x 1 columns]
```

le y est

```
847      MIXTURE/OTHER
447         TRUE/FALSE
2243        TRUE/FALSE
1202        TRUE/FALSE
1059     MIXTURE/OTHER
            ...
```

```
637      MIXTURE/OTHER
1022     MIXTURE/OTHER
1515        TRUE/FALSE
1432     MIXTURE/OTHER
628      MIXTURE/OTHER
Name: regrouped, Length: 1900, dtype: object

la taille de X_text est (1900, 1)
la taille de y_train est  (1900,)
les valeurs de TRUE et FALSE maintenant sont  MIXTURE/OTHER    950
TRUE/FALSE       950
Name: regrouped, dtype: int64
```

Etape 2 : Classification selon la colonne TITRE

Tests de Classification pour trouver le meilleur classifieur, les meilleurs paramètres de prétraitement et les meilleurs hyper_paramètres.

Avant de classifier on applique les entités nommées, Cette fonction utilise Spacy pour extraire les entités nommées à partir d'une liste de textes et les stocke dans une liste pour être affichées avec leurs labels correspondants.

```python
nlp = spacy.load("en_core_web_sm")
texte=dftrain['title']
#Créer une liste pour stocker les entités nommées
entities = []
# Traiter chaque texte individuellement et ajouter les entités à la
liste
for phrase in texte:
    phrase_str = str(phrase)  # Convertir l'élément en chaîne de
caractères
    doc = nlp(phrase_str)
    for ent in doc.ents:
        entities.append((ent.text, ent.label_))


# Afficher les entités nommées
for entity in entities:
    print(entity[0], entity[1])
```

```
Le flux de sortie a été tronqué et ne contient que les 5000 dernières
lignes.
Fox News ORG
Jeanine Show PERSON
Newton GPE
One CARDINAL
five CARDINAL
```

Britain GPE
less than 50 CARDINAL
Trump ORG
second Impeachment Trial ORG
the day DATE
House ORG
U.S. GPE
Utah GPE
3rd ORDINAL
Jason Chaffetz PERSON
the House Rules ORG
OAN ORG
Jack Posobiec PERSON
Donald Trump PERSON
second ORDINAL
Senate ORG
January 6 DATE
Congress ORG
62% PERCENT
Dan Jarvis PERSON
Winter DATE
the Iowa Caucuses ORG
Results Delayed PERSON
NHS ORG
two CARDINAL
each DAY DATE
Wisconsin GPE
Newly ORG
the Department of Instruction ORG
the past year DATE
Wisconsin GPE
2nd DATE
41st ORDINAL
more than half CARDINAL
today DATE
the Walker Administration ORG
Democratic Party of Wisconsin Chair ORG
Martha Laning PERSON
Thursday DATE
last year DATE
State ORG
Scott Walker PERSON
Wisconsin GPE
second ORDINAL
five days DATE
Walker's ORG
State ORG
the last six years DATE
36 CARDINAL
22.0 CARDINAL

20.0 CARDINAL
Wisconsin GPE
Midwest LOC
Kentucky GPE
41st ORDINAL
Wisconsin GPE
Wisconsin GPE
fourth ORDINAL
K-12 PRODUCT
2015-16 DATE
the Center on Budget and Policy Priorities ORG
Wisconsin GPE
Wisconsin GPE
Midwest LOC
ten CARDINAL
Wisconsin GPE
Republican NORP
Lisa Page PERSON
DNC Server Was Not Hacked ORG
Russia GPE
Competitive Elections Wisconsin ORG
State Assembly ORG
Only 3% PERCENT
Trump Voters ORG
Joe Biden PERSON
2000 years DATE
the Medieval Warm Period FAC
Little Ice Age LOC
JoNova ORG
Wasserman Schultz PERSON
2016 DATE
Senate ORG
2018 DATE
Overwhelmingly Benefitting Executives And Wealthy Shareholders ORG
The Washington Post ORG
Trump ORG
second Impeachment Trial ORG
the day DATE
House ORG
U.S. GPE
Utah GPE
3rd ORDINAL
Jason Chaffetz PERSON
the House Rules ORG
OAN ORG
Jack Posobiec PERSON
Donald Trump PERSON
second ORDINAL
Senate ORG
January 6 DATE

Congress ORG
the United States GPE
Rick Simpson PERSON
marijuana PERSON
Cannabis PERSON
Marijuana PERSON
Fox News Wins Pulitzer Prize for Reporting During Crisis — Daily Worl
ORG
DEMOCRATS NORP
The Great Barrier Reef WORK_OF_ART
2017 DATE
National Insurance ORG
Travis Rowley PERSON
Democrats NORP
CDC ORG
U.S. GPE
Tammy Baldwin PERSON
Wisconsin GPE
Ireland GPE
one CARDINAL
two CARDINAL
Congress ORG
Nancy Pelosi Retiring PERSON
2020 DATE
Didier Raoult PERSON
1,061 CARDINAL
91% PERCENT
Raoult PERSON
Hydroxychloroquine Azithromycin PERSON
COVID-19 ORG
0.5% PERCENT
first ORDINAL
Didier Raoult PERSON
French NORP
100% PERCENT
Hydroxychloroquine PERSON
Azithromycin PERSON
6 days DATE
80 CARDINAL
one 86 year-old DATE
74-year old DATE
one CARDINAL
the age of 86 DATE
today DATE
IHU Méditerranée Infection FAC
Tonight TIME
Raoult PERSON
OZ PERSON
Didier Raoult PERSON
1000 CARDINAL

99.3% PERCENT
1000 CARDINAL
20 CARDINAL
ICU ORG
7 CARDINAL
Raoult PERSON
Raoul PERSON
21:42 minutes TIME
one CARDINAL
Rastafarian Prez PERSON
Nana Akufo-Addo ORG
Trump ORG
thousands CARDINAL
Wisconsin GPE
Jeremy Corbyn Vows PERSON
One CARDINAL
Ten minutes TIME
each week DATE
Randi Weingarten PERSON
Sounds Off On Atlanta Cheating Scandal WORK_OF_ART
Jim Carrey PERSON
Christmas DATE
Arctic LOC
Trump ORG
Supreme Court ORG
Obamacare NORP
Devon PERSON
Chris King Statement PERSON
800 Percent PERCENT
Health Insurance Premiums for Low-Income Seniors ORG
Trumpcare ORG
Orlando GPE
Florida GPE
Tammy Baldwin PERSON
Wisconsin GPE
Election Year DATE
Kremlin ORG
U.S. GPE
US GPE
six weeks DATE
Coronavirus PERSON
Black Sea Fleet ORG
Russian NORP
Sea Breeze ORG
Lawrence O'Donnell' PERSON
Monday, September 16th, 2013 DATE
Earth LOC
15 years DATE
Conor Lamb PERSON
DeSantis Tells PERSON

Go Fuck Yourself" WORK_OF_ART
Trump ORG
Lin Wood Puts PERSON
Gov Brian Kemp PERSON
JUSTICE ORG
at least 40,000 CARDINAL
UK GPE
This Winter DATE
British NORP
100 years ago DATE
Up to 90% PERCENT
Covid PERSON
Laura Trott PERSON
Jason Kenny PERSON
Team GB's ORG
Olympics EVENT
Jerry Brown Blames Climate Change PERSON
Duncan Smith PERSON
National Review ORG
Noa Pothoven PERSON
Burnam Statement PERSON
GOP ORG
Vote Rejecting Medicaid Expansion ORG
Ohio Senate ORG
Rubio Hails PERSON
Bill To Defund PERSON
CDC ORG
Overnight TIME
Ohio Sets Up COVID-19 FEMA Camps ORG
Foreign Ministry ORG
Maria Zakharova PERSON
Moscow GPE
April 4, 2019 DATE
Russian NORP
Carnival Cruise Lines ORG
Detroit GPE
Voter Fraud PERSON
EU ORG
Nasa ORG
One CARDINAL
23,000 CARDINAL
Ballots PERSON
Biden Identified PERSON
Georgia GPE
Bernie PERSON
Biden PERSON
Scott Walker PERSON
Mary Burke PERSON
Paul Ryan PERSON
Seven CARDINAL

Radio Ad Campaign against Massive Immigration Increases ORG
Amnesty ORG
Israeli NORP
first ORDINAL
year DATE
Dan Aridor PERSON
Accelerated Evolution Biotechnologies Ltd. ORG
2000 DATE
ITEK GPE
SoAP ORG
day one DATE
a few weeks DATE
Aridor PERSON
Ilan Moran PERSON
Ilan Moran PERSON
an estimated 18.1 million MONEY
each year DATE
the International Agency for Research on Cancer ORG
sixth ORDINAL
second ORDINAL
second ORDINAL
Aridor ORG
Ilan Morad PERSON
SoAP ORG
2018 DATE
the Nobel Prize WORK_OF_ART
two CARDINAL
Morad GPE
first ORDINAL
Morad GPE
Morad GPE
Hanan Itzhaki PERSON
Morad GPE
at least three CARDINAL
Morad GPE
Morad GPE
three CARDINAL
three CARDINAL
Morad GPE
Morad PERSON
Morad PERSON
12 CARDINAL
Morad PERSON
one CARDINAL
two CARDINAL
Morad GPE
Today DATE
two CARDINAL
Morad GPE
only a few weeks DATE

Aridor GPE
Morad PERSON
first ORDINAL
a few years DATE
Christian NORP
One Year DATE
Republican NORP
James Lovelock PERSON
the end of this century DATE
One CARDINAL
Tetyana Obukhanych ORG
Ph.D. WORK_OF_ART
the Rockefeller University ORG
New York GPE
Harvard ORG
British Columbia GPE
Quebec GPE
2011 DATE
48% PERCENT
PhD WORK_OF_ART
one CARDINAL
two CARDINAL
a few years ago DATE
American NORP
first ORDINAL
two CARDINAL
11 DATE
Ukraine GPE
PhD WORK_OF_ART
Global Warming PERSON
Lancashire GPE
17th DATE
9 days DATE
Laura Trott PERSON
Jason Kenny PERSON
Team GB's ORG
Olympics EVENT
Capitol FAC
Legislature ORG
Pope PERSON
80 CARDINAL
Child Trafficking, Fraud — Conservative WORK_OF_ART
McFaul PERSON
Russian NORP
Putin PERSON
One CARDINAL
3 CARDINAL
Fifa ORG
England GPE
Scotland GPE

World Cup EVENT
Earth LOC
15 years DATE
NHS ORG
30% PERCENT
Trump ORG
thousands CARDINAL
Wisconsin GPE
Ariana Grande No Longer ORG
Japanese NORP
A decade DATE
Scottish National Party ORG
Hurricane Florence EVENT
Four CARDINAL
five CARDINAL
Syria GPE
EU ORG
100 Percent PERCENT
Donald Trump PERSON
CCP Virus ORG
American NORP
Trump PERSON
Antifa PERSON
the entire year DATE
Biden PERSON
Covid PERSON
at least 100 days DATE
CDC ORG
CDC ORG
Maricopa County GPE
Arizona Audit Update ORG
TENS OF THOUSANDS WORK_OF_ART
100,000 CARDINAL
The National File Reports ORG
CDC ORG
NATIONAL FILE ORG
CDC ORG
COVID ORG
at least 1,600 percent PERCENT
2020 DATE
the Public Health Initiative of the Institute for Pure ORG
COVID-19 Data Collection ORG
Comorbidity & Federal Law: A Historical Retrospective ORG
Henry Ealy PERSON
Michael McEvoy PERSON
Daniel Chong PERSON
John Nowicki PERSON
Monica Sava PERSON
Sandeep Gupta PERSON
David White PERSON

James Jordan PERSON
Daniel Simon PERSON
Paul Anderson PERSON
25 CARDINAL
over 100 CARDINAL
CDC ORG
Americans NORP
the World Health Organization ORG
PCR ORG
first ORDINAL
second ORDINAL
CDC ORG
Florida GPE
Orlando Health ORG
9.4 percent PERCENT
98 percent PERCENT
Gates ORG
Vaccine PERSON
America GPE
CDC ORG
a year DATE
2020 DATE
Covid PERSON
1600% PERCENT
CDC ORG
the Information Quality Act LAW
Paperwork Reduction Act LAW
Administrative Procedures Act ORG
(Publishing Journal — Institute for Pure ORG
Applied Knowledge / Public Health Policy Initiative ORG
CDC ORG
COVID-19 ORG
1,600% PERCENT
COVID-19 PERSON
All Concerned Citizens ORG
NATIONAL FILE ORG
the Institute for Pure ORG
Applied Knowledge ORG
CDC ORG
CDC ORG
CDC ORG
COVID-19 PERSON
All Concerned Citizens ORG
Covid PERSON
March 24th of 2020 DATE
decades DATE
March 24th DATE
CDC ORG
NVSS PRODUCT
COVID-19 Alert PERSON

2 CARDINAL
COVID-19 ORG
CDC ORG
2003 DATE
the previous 17 years DATE
COVID-19 ORG
COVID-19 ORG
2003 DATE
COVID ORG
COVID-19 ORG
as much as 1600% PERCENT
CDC ORG
2003 DATE
All Concerned Citizens ORG
Last spring DATE
CDC ORG
covid-19 DATE
April 14th DATE
CDC ORG
COVID-19 ORG
the Council of State and Territorial Epidemiologists ORG
April 5th DATE
CSTE ORG
Standardized ORG
2019 DATE
COVID-19 ORG
5 CARDINAL
CDC ORG
4 & 5 MONEY
5 CARDINAL
7 CARDINAL
All Concerned Citizens ORG
America GPE
The National File ORG
COVID-19 ORG
CDC ORG
the Institute for Health Metrics and Evaluation ORG
the Department of Justice ORG
US GPE
All Concerned Citizens WORK_OF_ART
last year DATE
Covid-19 hysteria PERSON
Western NORP
millions CARDINAL
Americans NORP
the Covid Cult ORG
America GPE
Bail Hearing PERSON
Ocala GPE
Florida GPE

30 Days DATE
Florida GPE
Abe Lincoln PERSON
Arrested Suspect Is A Black Lives Matter ORG
Amazon ORG
UN ORG
Vaccine PERSON
Africa - Global Research ORG
CLARKSBURG ORG
Jared Henry ORG
Maddie Mansell PERSON
Buckhannon Pike PERSON
Nutter Fort GPE
Mansell PERSON
Henry PERSON
Mansell PERSON
Mansell PERSON
Nikki Bowmar PERSON
the U.S. Small Business Administration ORG
Bowmar PERSON
approximately 114,391 CARDINAL
West Virginia GPE
99 percent PERCENT
280,213 CARDINAL
nearly 50 percent PERCENT
Bowmar PERSON
Mansell PERSON
first ORDINAL
Christ GPE
George "Gib" Brown PERSON
The James and Law Co. ORG
West Virginia GPE
nearly 30 CARDINAL
Clarksburg GPE
two CARDINAL
1903 DATE
Brown PERSON
James and Law ORG
Brown PERSON
the years DATE
Chamber ORG
the Robinson Grand Performing Arts Center ORG
Clarksburg GPE
Clarksburg GPE
years DATE
Brown PERSON
Dan Taylor PERSON
the West Virginia Community Development Hub's Energizing
Entrepreneurial Communities ORG
the last of three years DATE

four CARDINAL
Grafton ORG
Taylor PERSON
Taylor PERSON
Stephen Mann PERSON
Lewis County GPE
December DATE
Weston GPE
Mann ORG
two CARDINAL
one CARDINAL
401k PRODUCT
Mann PERSON
Manns PERSON
Lewis County GPE
Mann PERSON
Lewis County GPE
West Virginia GPE
Bowmar PERSON
Mann ORG
The Energizing Entrepreneurial Communities ORG
Grafton ORG
Taylor PERSON
All Aboard Grafton FAC
Taylor PERSON
Taylor County PERSON
first ORDINAL
February DATE
Patricia Henderson PERSON
the Taylor County Development Authority ORG
the Taylor County Commission ORG
Taylor County GPE
Weston GPE
Fairs and Festivals Committee ORG
Julia Spelsberg PERSON
Spelsberg PERSON
Weston GPE
North Central West Virginia LOC
The Small Business Administration ORG
Bowmar PERSON
USDA ORG
the Tax Cut FAC
Taylor PERSON
Nye County GPE
Change Wackos PERSON
California Court ORG
BANK BUSTING Labour's ORG
four CARDINAL
10billion MONEY
Central Europe LOC

Brexit PERSON
Britain GPE
Slovak GPE
David Cameron PERSON
ISIS ORG
Syria GPE
UK GPE
49,000 CARDINAL
Arctic LOC
Obama PERSON
November 14, 2014 DATE
China GPE
Canada GPE
Senate ORG
House ORG
three-degree QUANTITY
Warren Statement PERSON
Boeing ORG
Britain GPE
last seven years DATE
one CARDINAL
50 CARDINAL
Hillary Clinton Boards PERSON
The Climate Crisis Train To Nowhere ORG
UK GPE
Europe LOC
Spain GPE
France GPE
Germany GPE
Belgium GPE
the Netherlands GPE
Sweden GPE
Sunday DATE
Sweden GPE
100,000 CARDINAL
56.40 CARDINAL
U.K. GPE
69.60 CARDINAL
Spain GPE
60.88 CARDINAL
Italy GPE
58.16 CARDINAL
Sunday DATE
Johns Hopkins University ORG
U.K. GPE
fourth ORDINAL
Spain GPE
Italy GPE
sixth ORDINAL
eighth ORDINAL

Europe LOC
two CARDINAL
G-2101 CARDINAL
Pentagon ORG
2040 DATE
U.S. GPE
Jeff Merkley PERSON
Oregon GPE
first ORDINAL
James Lovelock PERSON
the end of this century DATE
Netizens ORG
ngayon ORG
China GPE
Exclusive: The TCF Center Election Fraud — Newly Discovered Video
Shows Late Night Deliveries WORK_OF_ART
Tens of Thousands CARDINAL
8 Hours TIME
Trump ORG
McCain ORG
Donald Trump PERSON
anti-American NORP
Kenosha GPE
Blake or Rittenhouse ORG
The Fight for Wisconsin's Soul - The New York Times ORG
China GPE
P4 PRODUCT
El Niño ORG
2015 DATE
Netflix GPE
Chris Watts PERSON
A&E ORG
Existence PERSON
Milwaukee GPE
fourth ORDINAL
Bible WORK_OF_ART
Andy Murray PERSON
John Inverdale PERSON
BBC ORG
Coronavirus Contains HIV Insertions ORG
European Court of Killers' Rights ORG
Antifa Provocateur PERSON
2017 DATE
650 CARDINAL
UK GPE
Joe Biden's Dept PERSON
Energy Blocked GPE
Texas GPE
Increasing Power Ahead of Killer Storm ORG
Brexit PERSON

£66bn DATE
NYS ORG
GOP ORG
Meets With PERSON
GOP ORG
IOC ORG
Beijing GPE
2022 DATE
June DATE
Earth LOC
European NORP
Mike Pence PERSON
Philadelphia GPE
Arctic LOC
Antarctic LOC
Democrats NORP
Americans NORP
NASA ORG
Arctic LOC
Pope PERSON
80 CARDINAL
Child Trafficking, Fraud — Conservative WORK_OF_ART
Two CARDINAL
Georgia GPE
Muhammad PERSON
Prophet Muhammad PERSON
the Religion of Islam ORG
Islamic NORP
MSF ORG
18 CARDINAL
first ORDINAL
Tammy Baldwin PERSON
Wisconsin GPE
Election Year DATE
2040 DATE
Evans PERSON
Dueker PERSON
St. Louis GPE
Congressional Budget Office ORG
House ORG
Democrats NORP
Remove National Anthem ORG
Sporting GPE
9 CARDINAL
The New York Times ORG
Up to 90% PERCENT
Covid PERSON
Emmett Till's Memorial Sign Shot PERSON
35 days DATE
Capitol ORG

Six CARDINAL
Congress ORG
MMR ORG
two CARDINAL
Wisconsin GPE
2018 DATE
Overwhelmingly Benefitting Executives And Wealthy Shareholders ORG
four million CARDINAL
a year DATE
last month DATE
JAMA GPE
monthly DATE
the American Medical Association ORG
COVID-19 ORG
15% PERCENT
John Aitken PERSON
San Antonio Fox 8 GPE
COVID-19 ORG
COVID-19 ORG
Joe Arpaio PERSON
America GPE
Toughest Sheriff PERSON
89 DATE
Tony Abbott PERSON
Rubio, Risch Statement On Senate Intelligence Committee Release Of
Interrogation Study ORG
Scott PERSON
1776 DATE
SIGNAL ORG
Paul Nuttall PERSON
Stoke LOC
Josh Mandel PERSON
The Washington Post ORG
Hayes PERSON
CDC Latest Update On Antibody Testing Results ORG
Kamala Harris PERSON
Superbowl Champions Refuse Invitation ORG
White House ORG
5 CARDINAL
American NORP
5 CARDINAL
almost two thirds CARDINAL
Britons GPE
second ORDINAL
Pfizer Covid PERSON
More than one CARDINAL
10 CARDINAL
50,000 CARDINAL
Chinese NORP
Olympics EVENT

Team GB PERSON
China GPE
Rio Olympics EVENT
Russia GPE
Scientific Data ORG
Hoax ORG
a Fake Pandemic WORK_OF_ART
500 MONEY
Brexit PERSON
Coca-Cola ORG
Vanilla Coke PERSON
Bernie PERSON
Biden PERSON
Cedar Rapids GPE
Article 50 LAW
Britain GPE
EU ORG
Texas GPE
This Labor Day DATE
Democrats NORP
Witney PERSON
David Cameron PERSON
EU ORG
first ORDINAL
Conservative NORP
British NORP
2015 DATE
year DATE
British NORP
a year later DATE
Britain GPE
EU ORG
Britain GPE
EU ORG
Cameron PERSON
Cameron ORG
Parliament ORG
544 CARDINAL
Second Reading ORG
seven CARDINAL
Dems NORP
eight CARDINAL
2015 DATE
Dems NORP
Cameron PERSON
Dems NORP
EU ORG
years DATE
the Conservative Party ORG
2008 DATE

Twitter PRODUCT
Conservative NORP
Nick Clegg PERSON
Britain GPE
the European Union ORG
thirty years DATE
British NORP
Britain GPE
the European Union ORG
Liberal Democrats NORP
Europe LOC
Britain GPE
EU ORG
2008 DATE
the year DATE
Clegg GPE
the House of Commons ORG
Michael Martin PERSON
EU ORG
Ed Davey's PERSON
Clegg PERSON
Ed Davey PERSON
today DATE
EU ORG
Davey PERSON
Dems NORP
Will the Chair WORK_OF_ART
Democrat NORP
Britain GPE
EU ORG
House ORG
Lib Dem PERSON
Jo Swinson PERSON
the weekend DATE
Swinson PERSON
The Liberal Democrats WORK_OF_ART
Europe LOC
EU ORG
House ORG
Clegg GPE
Swinson PERSON
EU ORG
EU ORG
House ORG
The Lib Dems WORK_OF_ART
2010 DATE
General Election ORG
a Coalition Government ORG
Conservatives NORP
British NORP

UK GPE
EU ORG
Dems NORP
British NORP
EU ORG
first ORDINAL
Brexit PERSON
second ORDINAL
second ORDINAL
Remain ORG
Lib Dem PERSON
first ORDINAL
Dems NORP
Norman Lamb PERSON
Dems NORP
first ORDINAL
Parliament ORG
democrats NORP
British NORP
Republican NORP
Cedar Rapids GPE
EU ORG
Judy Murray's PERSON
Scottish NORP
London GPE
Brussels GPE
Boko Haram PERSON
Shekau PERSON
Myanmar GPE
JK Rowling Confirms Stance Against Transgender Women ORG
over 200,000 CARDINAL
NHS ORG
2010 DATE
£330 million MONEY
Democrats NORP
RESPONDS ORG
Chu Decries PERSON
Belgium GPE
3 CARDINAL
Almost 2,400 CARDINAL
TWO WEEKS DATE
Trump ORG
second Impeachment Trial ORG
the day DATE
House ORG
U.S. GPE
Utah GPE
3rd ORDINAL
Jason Chaffetz PERSON
the House Rules ORG

OAN ORG
Jack Posobiec PERSON
Donald Trump PERSON
second ORDINAL
Senate ORG
January 6 DATE
Congress ORG
Tories NORP
Brexit PERSON
NHS ORG
Russian NORP
WADA GPE
Moscow GPE
IOC ORG
Olympics EVENT
Joe Biden PERSON
the United States GPE
Rick Simpson PERSON
marijuana PERSON
Cannabis PERSON
Marijuana PERSON
Republican Party of Wisconsin Launches Radio ORG
Ads Against PERSON
Four CARDINAL
Capitol FAC
Biden PERSON
Friends Busted Partying and Singing Without Masks or Social Distancing
ORG
£30 billion MONEY
NHS ORG
JOINS ORG
GEORGIA GPE
NEVADA GPE
TRUMP GPE
post-Brexit NORP
recent months DATE
Adam Castillejo PERSON
more than 30 months DATE
Lancet ORG
Mr Castillejo PERSON
2011 DATE
Timothy Brown PERSON
the "Berlin Patient ORG
first ORDINAL
three and half years DATE
Adam Castillejo - the PERSON
40-year-old DATE
London Patient WORK_OF_ART
a year DATE
first ORDINAL

Prof Ravindra PERSON
Kumar Gupta PERSON
the University of Cambridge ORG
BBC News ORG
two and a half years DATE
first ORDINAL
nine years ago DATE
the Berlin Patient FAC
millions CARDINAL
Prof Gupta PERSON
HIV-1 CARDINAL
two CARDINAL
CCR5 ORG
CCR5 ORG
Chinese NORP
He Jiankui PERSON
first ORDINAL
99% PERCENT
Castillejo PERSON
Mr Brown PERSON
Castillejo PERSON
the New York Times ORG
Sharon Lewin PERSON
the University of Melbourne ORG
Australia GPE
the London Patient ORG
Heathrow FAC
third ORDINAL
UK GPE
Yahaya Bello PERSON
Hal Turner PERSON
50,000 CARDINAL
Chinese NORP
Maine GPE
Henry Kissinger PERSON
the New World Order ORG
Brexit PERSON
75 CARDINAL
MPH ORG
32.2bn MONEY
just ONE year DATE
a quarter DATE
just five years DATE
Antifa PERSON
Trump ORG
Missouri GPE
Las Vegas GPE
Teen PERSON
FDA ORG
Juuls PERSON

Beef WORK_OF_ART
Bloomberg PERSON
2000 years DATE
the Medieval Warm Period FAC
Little Ice Age LOC
JoNova ORG
White House ORG
Trump ORG
Iran GPE
Erbil ORG
Iraq GPE
Historical Data ORG
Global Warming ORG
Fastest Rate FAC
Last 28 Centuries DATE
Swedish NORP
Photos Flood Social ORG
UK GPE
European Union ORG
British NORP
50/50 CARDINAL
EU ORG
2055 CARDINAL
Vladimir Putin PERSON
Europe LOC
Capitol ORG
Six CARDINAL
Congress ORG
Iowa GPE
Hurricane Florence EVENT
30% PERCENT
One CARDINAL
Tilapia ORG
Tilapia ORG
Wake Forest University School of Medicine ORG
Tilapia PERSON
Tilapia ORG
Tilapia ORG
one CARDINAL
Tilapia PERSON
Tilapia ORG
Tilapia PERSON
Tilapia PERSON
Tilapia GPE
Italy Arrests Cardinals And Seizes ORG
Pro-Life Groups Upset ORG
UN ORG
Ecuador GPE
Antifa PERSON
Trump ORG

£30 billion MONEY
NHS ORG
DAILY GPE
Court ORG
Ameen Jogee ORG
Justine Greening PERSON
China GPE
the World Health Organization ORG
the U.S. National Institutes of Health ORG
more than 2 million CARDINAL
Wuhan GPE
China GPE
the United States GPE
COVID-19 ORG
Fauci PERSON
Coronavirus PERSON
Art Moore   PERSON
February 1, 2021 DATE
Tweet Gab Share PERSON
Anthony Fauci PERSON
National Institute of Allergy and Infectious Diseases ORG
Thursday, July 30, 2020 DATE
Washington GPE
D.C. GPE
White House ORG
Tia Dufour PERSON
Feb. 8, 2021: DATE
USA Today ORG
March DATE
April 2020 DATE
Chinese NORP
Fadela Chaib PERSON
The Scripps Research Institute ORG
FactCheck.org ORG
Feb. 7, 2020 DATE
the last 20 years DATE
Timothy Sheahan PERSON
the University of North Carolina ORG
Chapel Hill FAC
96% PERCENT
China GPE
the World Health Organization ORG
the U.S. National Institutes of Health ORG
more than 2 million CARDINAL
Wuhan GPE
China GPE
the United States GPE
Newsweek ORG
April 2020 DATE
NIH ORG

Anthony Fauci PERSON
more than 200 CARDINAL
Obama PERSON
2014 DATE
U.S. GPE
Fauci ORG
2017 DATE
April 2020 DATE
the Wuhan Institute of Virology ORG
China GPE
Fauci's National Institute of Allergy and Infectious Diseases ORG
Steve Hilton PERSON
Fox News ORG
The Next Revolution WORK_OF_ART
a decade ago DATE
Yunnan province GPE
Wuhan GPE
1,000 miles QUANTITY
Hilton PERSON
Sunday DATE
Wuhan GPE
February 2020 DATE
Yunnan GPE
COVID-19 PERSON
Hilton PERSON
NIH ORG
GenBank ORG
only one CARDINAL
COVID-19 ORG
a decade ago DATE
Yunnan GPE
Hilton GPE
Wuhan GPE
Yunnan GPE
GenBank ORG
Yunnan GPE
Wuhan GPE
Hilton PERSON
two CARDINAL
two CARDINAL
COVID-19 PERSON
NAID ORG
a Nov. 30, 2017 DATE
NAID ORG
Hilton PERSON
a decade ago DATE
NAID ORG
Wuhan GPE
first ORDINAL
COVID-19 ORG

the fall of 2019 DATE
IOC ORG
Beijing GPE
2022 DATE
Masari NORP
MSF ORG
18 CARDINAL
first ORDINAL
Apocalyptic Sea-Level Rise ORG
1bn MONEY
next two years DATE
Europe LOC
Colin Powell PERSON
Michael Flynn PERSON
Central Park's FAC
COVID-19 FAC
The Prices Of Food Items Are Coming Down ORG
Garba Shehu PERSON
2005 DATE
six months DATE
two-year-old DATE
months DATE
COVID ORG
one CARDINAL
about 6-8 weeks CARDINAL
October 23 DATE
MIT ORG
the University of California San Francisco ORG
Decrease in Hospitalizations for COVID-19 WORK_OF_ART
Mask Mandates PERSON
1083 DATE
U.S. GPE
November 4 DATE
months DATE
COVID ORG
one CARDINAL
about 6-8 weeks CARDINAL
October 23 DATE
MIT ORG
the University of California San Francisco ORG
Decrease in Hospitalizations for COVID-19 WORK_OF_ART
Mask Mandates PERSON
1083 DATE
U.S. GPE
November 4 DATE
second ORDINAL
third ORDINAL
Danish NORP
months DATE
Last week DATE

the Naval Medical Research Center ORG
Maryland GPE
Mount Sinai Hospital FAC
New York GPE
the New England Journal of Medicine ORG
U.S. Marine Corps ORG
Parris Island GPE
over 3,000 CARDINAL
two CARDINAL
day 1 DATE
day 14 DATE
2.8% PERCENT
1.7% PERCENT
at least 6 feet QUANTITY
daily DATE
Six CARDINAL
8-hour TIME
Fauci PERSON
Marine NORP
Army ORG
Fort Benning FAC
90% PERCENT
Trump Revokes All ORG
First ORDINAL
Marine Corps ORG
Three-quarters CARDINAL
NHS ORG
overnight TIME
John Roberts PERSON
Island GPE
last month DATE
JAMA GPE
monthly DATE
the American Medical Association ORG
COVID-19 ORG
15% PERCENT
John Aitken PERSON
San Antonio Fox 8 GPE
COVID-19 ORG
COVID-19 ORG
Putin PERSON
Democrats NORP
CIA ORG
IOC ORG
Beijing GPE
2022 DATE
AUTOPSY ORG
Chadwick Boseman PERSON
Toronto GPE
Today DATE

NIPOST ORG
Osogbo GPE
Detroit GPE
Voter Fraud PERSON
David Cameron PERSON
ISIS ORG
Syria GPE
London GPE
Greenwich GPE
Lambeth PERSON
Foreign Ministry ORG
US GPE
anti-Russian NORP
Skripal PERSON
One CARDINAL
four CARDINAL
25 years DATE
Newton GPE
Jerry Brown Blames Climate Change PERSON
Olympics EVENT
Team GB PERSON
China GPE
Rio Olympics EVENT
Scott Walker's PERSON
NHS ORG
thousands CARDINAL
Olympics EVENT
Team GB PERSON
China GPE
Rio Olympics EVENT
Fire Expert Who PERSON
Capitol ORG
Six CARDINAL
Congress ORG
Capitol Shooting Exposed ORG
Hollywood GPE
Tom Graves Defends PERSON
the Referendum Over ORG
British NORP
Joe Biden's Dept PERSON
Energy Blocked GPE
Texas GPE
Increasing Power Ahead of Killer Storm ORG
Biden Calls PERSON
Done Meeting PERSON
NASA ORG
David McKinley PERSON
the White House Arresting Congress FAC
Marco Rubio Just Publicly Shamed His Own Constituents For Flocking
PERSON

Wisconsin GPE
Newly ORG
the Department of Instruction ORG
the past year DATE
Wisconsin GPE
2nd DATE
41st ORDINAL
more than half CARDINAL
today DATE
the Walker Administration ORG
Democratic Party of Wisconsin Chair ORG
Martha Laning PERSON
Thursday DATE
last year DATE
State ORG
Scott Walker PERSON
Wisconsin GPE
second ORDINAL
five days DATE
Walker's ORG
State ORG
the last six years DATE
36 CARDINAL
22.0 CARDINAL
20.0 CARDINAL
Wisconsin GPE
Midwest LOC
Kentucky GPE
41st ORDINAL
Wisconsin GPE
Wisconsin GPE
fourth ORDINAL
K-12 PRODUCT
2015-16 DATE
the Center on Budget and Policy Priorities ORG
Wisconsin GPE
Wisconsin GPE
Midwest LOC
ten CARDINAL
Wisconsin GPE
Republican NORP
Nolte PERSON
this week DATE
Scottish National Party ORG
Boko Haram PERSON
Shekau PERSON
200 CARDINAL
NHS ORG
Schools ORG
1.3bn MONEY

the U.S. ORG
Amy Willis PERSON
US GPE
The National Cancer Institute ORG
the US Department of Health ORG
April this year DATE
US GPE
National Institute on Drug Abuse ORG
the United States GPE
George Osborne's PERSON
2015 DATE
SUPREME COURT OF ORG
THE UNITED STATES GPE
18A335 CARDINAL
RICHARD BRAKEBILL PERSON
ET AL ORG
U.S. GPE
Twins ORG
Iowa GPE
Brussels GPE
Britain GPE
36 MONEY
Army ORG
Massachusetts GPE
Germany GPE
US GPE
CDC ORG
Brexit Britain PERSON
EU ORG
SURGES ORG
Historical Data ORG
Global Warming ORG
White Men Have Committed More Mass Shootings ORG
Any Other Group ORG
1984 DATE
the New York Times ORG
Trump ORG
Pfizer PERSON
UN Supports Disinformation Campaign ORG
Harris PERSON
Biden PERSON
Ban Fracking PERSON
Kirk Douglas Dead PERSON
4 Days DATE
The Phony War Against CO2
 ORG
Denzel Washington Praises Police, Denounces BLM ORG
State ORG
Georgia GPE
Trump ORG

USC ORG
Joel Hay PERSON
Miranda ORG
NHS ORG
thousands CARDINAL
Rhode Island General Assembly ORG
USA GPE
World Cup EVENT
Netherlands GPE
2 CARDINAL
Joy Covey PERSON
Amazon ORG
50 CARDINAL
Woodside GPE
Joy Covey PERSON
Amazon ORG
50 CARDINAL
Woodside GPE
Australia GPE
Great Barrier Reef WORK_OF_ART
North Carolina's GPE
Thousands CARDINAL
New Jersey Sept GPE
14, 2020 DATE
Videos GPE
UFO Sighting News ORG
FG ORG
Scottish Parliament ORG
Tories NORP
Thousands CARDINAL
Erie GPE
Pennsylvania GPE
MSM ORG
Italy GPE
Nazi NORP
pro-Russian NORP
Sarah Parker PERSON
Supreme Court ORG
051316 DATE
Bill Withers PERSON
$2 Million MONEY
Trump Re-election Campaign ORG
Florida GPE
GOP ORG
Mitt Romney PERSON
$1 million MONEY
Senate ORG
Utah GPE
Paid Liberal Trolls of America ORG
6 hours TIME

30 MONEY
40 MONEY
70 cents MONEY
1 MONEY
Soros PERSON
Joe Barron PERSON
George Soros PERSON
500 MONEY
Art Tubolls PERSON
163 CARDINAL
Soros PERSON
Deluth GPE
Charlie Kirk PERSON
Turning Point PERSON
NASCAR ORG
MAGA Hats PERSON
DISGRACEFUL ORG
EU ORG
British NORP
140MILLION MONEY
Republican Party of Wisconsin Launches Radio ORG
Ads Against PERSON
Four CARDINAL
FEMA ORG
Smart" Guillotines Placed WORK_OF_ART
FEMA Internment Camps ORG
NHS ORG
Japan GPE
15-year DATE
under 5,000 DATE
the COVID-19 Relief Fund ORG
2030 DATE
Marijuana PERSON
Meghan Markle PERSON
America GPE
first ORDINAL
95% PERCENT
Joe Biden PERSON
FDA Shocking Study: Cells Used ORG
Dan Jarvis PERSON
Winter DATE
Next year DATE
the year DATE
Arctic LOC
Miles PERSON
CDC ORG
the American Academy of Pediatrics ORG
Andrew Zimmerman PERSON
Kennedy Krieger Institute ORG
Johns Hopkins University ORG

third ORDINAL
years DATE
Zimmerman PERSON
Zimmerman PERSON
more than five thousand CARDINAL
2009 DATE
the late 1990s DATE
the Kennedy Krieger Institute ORG
Baltimore GPE
Zimmerman PERSON
Jon Poling PERSON
2000 DATE
nineteen-month-old DATE
Hannah PERSON
Hannah PERSON
Zimmerman PERSON
Richard Kelley PERSON
Kennedy Krieger's PERSON
Kennedy Krieger PERSON
Hannah PERSON
Georgetown GPE
MD GPE
Teri PERSON
Hannah PERSON
the next few years DATE
Zimmerman PERSON
2008 DATE
Hannah Poling PERSON
Zimmerman PERSON
Zimmerman PERSON
Zimmerman PERSON
Hannah PERSON
CDC ORG
the Department of Justice ORG
twenty-million-dollar MONEY
Polings FAC
2010 DATE
Hannah PERSON
Hannah Poling PERSON
Zimmerman PERSON
Kelley PERSON
the Department of Justice ORG
Drs PERSON
Zimmerman PERSON
Kelley PERSON
first ORDINAL
more than thirty years DATE
Tennessee GPE
Yates Hazlehurst PERSON
one CARDINAL

three CARDINAL
2009 DATE
only a year DATE
Poling ORG
Zimmerman.3 Drs PERSON
Zimmerman PERSON
Kelley PERSON
Hazlehurst NORP
Yates GPE
2009 DATE
the Omnibus Autism Proceeding ORG
Zimmerman PERSON
more than five thousand CARDINAL
2017 DATE
Drs PERSON
Kelley PERSON
Zimmerman PERSON
Yates GPE
The "Vaccine Court ORG
the National Childhood Vaccine Injury Act ORG
1986 DATE
the United States Court of Federal Claims ORG
American NORP
the Department of Health and Human Services ORG
HHS ORG
Department of Justice ORG
1989 DATE
more than $3.8 billion MONEY
Americans NORP
1998 DATE
twelve hundred CARDINAL
eighteen thousand CARDINAL
DTP ORG
MMR ORG
second ORDINAL
approximately 34 percent PERCENT
2017 DATE
just over $282 million MONEY
Rolf Hazlehurst PERSON
Tennessee GPE
Yates PERSON
the US Congress ORG
2013 DATE
Rolf Hazlehurst ORG
Vaccine PERSON
Discovery PRODUCT
American NORP
Clean Energy and Pollution Reduction Act ORG
2015 DATE
200,000 MONEY

Wisconsin GPE
Miraculous Vanishing Ballot Rejection Rate ORG
The Great Barrier Reef WORK_OF_ART
Abe Lincoln PERSON
Arrested Suspect Is A Black Lives Matter ORG
OMINOUS ORG
U.S. GPE
Mueller ORG
Russia GPE
more than $25 million MONEY
Democrats NORP
Americans NORP
Scott Walker PERSON
$1 million MONEY
The Squad Wants Bible Deemed To Be 'Hate Speech ORG
Senate ORG
Republican NORP
Democrats NORP
21 Million CARDINAL
China GPE
Coronavirus Outbreak GPE
Iowa GPE
U.K. GPE
Chu Decries PERSON
London GPE
Obama ORG
Myanmar GPE
Muhammad PERSON
Prophet Muhammad PERSON
the Religion of Islam ORG
Islamic NORP
Joe Biden Tells PERSON
Joe Biden PERSON
Labour ORG
Fulton County Blocks Credentialed ORG
Republican NORP
Bill PERSON
Iran GPE
Delta Force Raids Biden Compound ORG
Ukraine GPE
$2.2 Billion MONEY
Social Security ORG
Cuomo's Budget ORG
A&E ORG
Romney Endorses PERSON
Joe Biden PERSON
America GPE
Zamfara PERSON
Trump Files PERSON
Georgia GPE

State ORG
Leaking Confidential Litigation Call ORG
May 2018 DATE
fourth ORDINAL
the month of May DATE
NOAA ORG
1880 DATE
March-May DATE
fourth ORDINAL
monthly DATE
NOAA ORG
National Centers for Environmental Information ORG
NOAA ORG
May 2018 DATE
May DATE
1.44 CARDINAL
F PRODUCT
the 20th century DATE
58.6 CARDINAL
F PRODUCT
fourth ORDINAL
May DATE
1880-2018 DATE
years 2014-2018 DATE
five CARDINAL
2016 CARDINAL
May DATE
1.58 CARDINAL
May 2018 DATE
the 42nd consecutive May DATE
the 401st consecutive month DATE
the 20th century DATE
North America LOC
Europe LOC
Asia LOC
Atlantic LOC
Pacific LOC
Indian Oceans NORP
May DATE
Canada GPE
Atlantic Ocean LOC
Greenland GPE
May DATE
2.05 CARDINAL
the 20th century DATE
52.0 CARDINAL
F PRODUCT
May DATE
2011 DATE
2013 DATE

seventh ORDINAL
May DATE
139-year DATE
May 2018 DATE
U.S. GPE
Europe LOC
5.4 CARDINAL
Canada GPE
Russia GPE
5.4 CARDINAL
Europe LOC
May DATE
continental ORG
1910 DATE
4.97 CARDINAL
2003 DATE
1.66 CARDINAL
F PRODUCT
May 2018 DATE
first ORDINAL
May DATE
3.6 CARDINAL
F PRODUCT
2.0 CARDINAL
European NORP
May DATE
Asia LOC
May DATE
2009 DATE
May DATE
1.19 CARDINAL
the 20th century DATE
monthly DATE
61.3 CARDINAL
fourth ORDINAL
May DATE
years 2014-2018 DATE
five CARDINAL
2016 CARDINAL
May DATE
1.39 CARDINAL
May 2018 DATE
May DATE
Arctic LOC
second ORDINAL
40-year DATE
420,000 square miles QUANTITY
8.1 percent PERCENT
1981-2010 DATE
the National Snow and Ice Data Center(link ORG

NOAA ORG
NASA ORG
May Arctic LOC
2016 DATE
Arctic LOC
Bering GPE
Chukchi Seas PERSON
May DATE
330,000 square miles QUANTITY
8.6 percent PERCENT
1981-2010 DATE
third ORDINAL
May DATE
Antarctic sea LOC
May DATE
the Weddell Sea LOC
Seasonal DATE
March-May 2018 DATE
March-May DATE
1.48 CARDINAL
the 20th century DATE
56.7 CARDINAL
F PRODUCT
fourth ORDINAL
March-May DATE
1880-2018 DATE
0.45 CARDINAL
F cooler PRODUCT
2016 DATE
March-May DATE
2014 DATE
three-month DATE
Europe LOC
the Middle East LOC
Africa LOC
China GPE
Mongolia GPE
South America LOC
Atlantic LOC
Indian NORP
Pacific Oceans NORP
March-May 2018 DATE
March-May DATE
2.34 CARDINAL
the 20th century DATE
46.4 CARDINAL
F PRODUCT
fifth ORDINAL
March-May DATE
139-year DATE

Five CARDINAL
March-May 2018 DATE
nine CARDINAL
continental ORG
1910 DATE
Europe LOC
South America LOC
third ORDINAL
March-May DATE
March-May DATE
1.15 CARDINAL
the 20th century DATE
61.0 CARDINAL
F PRODUCT
fourth ORDINAL
March-May DATE
Year-to-Date DATE
January-May 2018 DATE
1.39 CARDINAL
the 20th century DATE
55.5 CARDINAL
F PRODUCT
2010 DATE
fourth ORDINAL
January-May DATE
139-year DATE
2018 year DATE
0.59 CARDINAL
2016 DATE
2.18 CARDINAL
the 20th century DATE
42.8 CARDINAL
F PRODUCT
2002 DATE
fifth ORDINAL
January-May DATE
1.08 CARDINAL
the 20th century DATE
60.8 CARDINAL
fifth ORDINAL
January-May DATE
1880-2018 DATE
Noa Pothoven PERSON
Antonio Villaraigosa PERSON
Gavin Newsom PERSON
one million CARDINAL
Birth Certificates PERSON
Federal Bank ORG
US GPE
50,000 MONEY

ACCLAIMED ISRAELI ASTROPHYSICIST ORG
Austin PERSON
BRAZIL GPE
AstraZeneca ORG
Oxford University ORG
Anvisa ORG
one CARDINAL
three CARDINAL
Brazil GPE
Oxford University ORG
the UK Government ORG
100 million CARDINAL
US GPE
September 6 DATE
India GPE
South Africa GPE
AstraZeneca ORG
Europe LOC
Inclusive Vaccine Alliance ORG
400 million CARDINAL
European NORP
UK GPE
30 million CARDINAL
BioNtech/Pfizer ORG
60 million CARDINAL
Valneva - in PERSON
six CARDINAL
UK GPE
Patrick Vallance PERSON
next spring DATE
the Commons and Lords National Security Strategy Committee ORG
AstraZeneca ORG
the turn of the year DATE
UK GPE
Vaccine Taskforce FAC
Kate Bingham PERSON
UK GPE
four CARDINAL
six CARDINAL
three CARDINAL
one CARDINAL
two CARDINAL
Covid PERSON
26,688 CARDINAL
UK GPE
789,229 CARDINAL
191 CARDINAL
today DATE
Paul Ryan PERSON
The New York Times ORG

Joe Arpaio PERSON
America GPE
Toughest Sheriff PERSON
89 DATE
Russia GPE
Trump ORG
2020 DATE
five years DATE
Tlaib Files Lawsuit ORG
The New York Times ORG
Florida GPE
GOP ORG
Pentagon ORG
Buhari To Resign PERSON
May 2023 DATE
Lai Mohammed PERSON
Obama PERSON
Cameron PERSON
the United Kingdom GPE
annual DATE
American NORP
the Department of Justice ORG
the U.S. Vaccine Court ORG
three months DATE
Advisory Commission ORG
Childhood Vaccines ORG
the U.S. Department of Health and Human Services ORG
millions of dollars MONEY
Health Impact News ORG
December 6, DATE
2018 DATE
Catharine Reeves PERSON
144 CARDINAL
the past 3 months DATE
89 CARDINAL
more than 5 years DATE
89 CARDINAL
77 CARDINAL
one CARDINAL
the United States GPE
GBS — Guillain–Barré ORG
one CARDINAL
SIRVA ORG
Vaccine Administration ORG
Adhesive Capsulitis PERSON
Frozen Shoulder Syndrome PERSON
Myasthenia GPE
Peripheral Neuropathy PERSON
five years DATE
Berger PERSON

NC GPE
Netflix GPE
Chris Watts PERSON
AG ORG
2 CARDINAL
Milwaukee GPE
NHS ORG
Florida GPE
Hank Aaron PERSON
Moderna COVID GPE
Jan. 5 DATE
the New York Times ORG
Atlanta Braves ORG
86-year-old DATE
Hall of Famer's ORG
today DATE
CNN ORG
Aaron PERSON
Aaron PERSON
earlier this month DATE
Moderna PERSON
the Associated Press ORG
Aaron PERSON
the Morehouse School of Medicine ORG
Atlanta GPE
Black Americans NORP
AP ORG
first ORDINAL
two CARDINAL
the Morehouse School of Medicine ORG
African Americans NORP
Tuskegee ORG
U.S. GPE
Black LOC
December DATE
VOX ORG
COVID ORG
Germany GPE
Norway GPE
43 CARDINAL
Pfizer PERSON
China GPE
Aaron PERSON
Robert F. Kennedy PERSON
Jr. PERSON
Children's Health Defense ORG
Aaron PERSON
COVID ORG
Kennedy PERSON
the Vaccine Adverse Events Reporting System ORG

the U.S. Department of Health and Human Services ORG
HHS ORG
2001 DATE
HHS ORG
fewer than 1% PERCENT
last week DATE
California GPE
330,000 CARDINAL
Moderna PERSON
Defender PERSON
56-year-old DATE
Florida GPE
about two weeks DATE
first ORDINAL
Pfizer ORG
Florida GPE
Pfizer ORG
Ohio GPE
Florida GPE
Rick Scott PERSON
Paris Climate Agreement ORG
Americans NORP
One CARDINAL
Tetyana Obukhanych ORG
Ph.D. WORK_OF_ART
the Rockefeller University ORG
New York GPE
Harvard ORG
British Columbia GPE
Quebec GPE
2011 DATE
48% PERCENT
PhD WORK_OF_ART
one CARDINAL
two CARDINAL
a few years ago DATE
American NORP
first ORDINAL
two CARDINAL
11 DATE
Ukraine GPE
PhD WORK_OF_ART
Black Sea Fleet ORG
Russian NORP
Sea Breeze ORG
tens of thousands CARDINAL
The Ron Paul Institute for Peace and Prosperity : The Coronavirus Hoax
ORG
200,000 MONEY
Democratic NORP

Southern States GPE
California GPE
the Changing Climate - The New York Times ORG
800,000 years DATE
The Unvaccinated Amish Rarely Get Cancer ORG
78 CARDINAL
Epstein Isle LOC
Ted Cruz PERSON
years DATE
Two-Thirds CARDINAL
Canada GPE
at least 13 CARDINAL
COVID-19 ORG
the United States of America GPE
the past few weeks DATE
COVID-19 ORG
COVID-19 ORG
Donald Trump PERSON
Canada GPE
800 CARDINAL
Cannabis PRODUCT
COVID-19 ORG
13 CARDINAL
ACE2 PRODUCT
C. NORP
TMPRSS2 PRODUCT
COVID-19 PERSON
13 CARDINAL
Canada's CARDINAL
Trump ORG
Harvard Medical School ORG
Mandeep Mehra PERSON
more than 96,000 CARDINAL
15,000 CARDINAL
the Washington Post ORG
34 percent PERCENT
137 percent PERCENT
Trump ORG
45 percent PERCENT
411 percent PERCENT
37 percent PERCENT
256 percent PERCENT
37 percent PERCENT
301 percent PERCENT
Donald Trump PERSON
US GPE
COVID-19 ORG
Plaquenil NORP
COVID-19 PERSON
COVID-19 ORG

PM May DATE
Sajd Javid PERSON
Adam Castillejo PERSON
more than 30 months DATE
Lancet ORG
Mr Castillejo PERSON
2011 DATE
Timothy Brown PERSON
the "Berlin Patient ORG
first ORDINAL
three and half years DATE
Adam Castillejo - the PERSON
40-year-old DATE
London Patient WORK_OF_ART
a year DATE
first ORDINAL
Prof Ravindra PERSON
Kumar Gupta PERSON
the University of Cambridge ORG
BBC News ORG
two and a half years DATE
first ORDINAL
nine years ago DATE
the Berlin Patient FAC
millions CARDINAL
Prof Gupta PERSON
HIV-1 CARDINAL
two CARDINAL
CCR5 ORG
CCR5 ORG
Chinese NORP
He Jiankui PERSON
first ORDINAL
99% PERCENT
Castillejo PERSON
Mr Brown PERSON
Castillejo PERSON
the New York Times ORG
Sharon Lewin PERSON
the University of Melbourne ORG
Australia GPE
the London Patient ORG
1984 DATE
the New York Times ORG
Trump ORG
Texans NORP
99.96% PERCENT
two CARDINAL
Pfizer ORG
Covid PERSON

Israel GPE
U.S. GPE
COVID ORG
Texas GPE
Julia Roberts: 'Michelle Obama PERSON
Toilet ORG
City Council ORG
Up to a third CARDINAL
Texans NORP
Honduran NORP
Florida GPE
Rick Scott PERSON
Paris Climate Agreement ORG
Americans NORP
11-year-olds ORDINAL
11-year-olds ORDINAL
The Washington Post ORG
1% PERCENT
US GPE
BOMBSHELL ORG
440% PERCENT
the coming flu season DATE
British Government Study Confirms Covid-19 Vaccine Risk: Infections
ORG
Fortnight GPE
January DATE
80 CARDINAL
Pemberley House ORG
Basingstoke GPE
England GPE
a third CARDINAL
24 CARDINAL
Germany GPE
British NORP
Gibraltar GPE
Pfizer ORG
16th January DATE
33 thousand CARDINAL
2761 CARDINAL
9 CARDINAL
December DATE
almost a hundred CARDINAL
The United Arab Emirates GPE
Chinese NORP
Sinopharm ORG
Pfizer ORG
December DATE
Amer Sharif PERSON
the Covid-19 Command & Control Centre ORG
Dubai GPE

UAE ORG
first ORDINAL
100% PERCENT
10th February DATE
daily DATE
18 CARDINAL
first ORDINAL
last spring DATE
Israel GPE
19th December DATE
New York Times ORG
Alex Berenson PERSON
Israelis NORP
January DATE
Palestine GPE
December DATE
Palestinians NORP
Israeli NORP
Hervé Seligmann PERSON
Aix-Marseilles University ORG
about 40 CARDINAL
the three weeks DATE
first ORDINAL
second ORDINAL
Israel GPE
Pfizer ORG
the first week DATE
Raine PERSON
UK GPE
MHRA ORG
two CARDINAL
first ORDINAL
The Public Health England ORG
Matt Hancock PERSON
80% PERCENT
48% PERCENT
first ORDINAL
Pfizer ORG
Astra ORG
Zeneca NORP
Mary Ramsay PERSON
PHE ORG
hundreds CARDINAL
British NORP
the Bill & Melinda Gates Foundation LAW
Covid PERSON
Covid PERSON
FEMA ORG
Smart" Guillotines Placed WORK_OF_ART
FEMA Internment Camps ORG

Dewhurst Campaign Owes Vendors ORG
Nearly $1 Million MONEY
Arctic LOC
Winter Sea Ice Drops LOC
The American Medical Association ORG
the past year DATE
Just a few days DATE
the past year DATE
AMA ORG
only 12 percent PERCENT
AMA ORG
AMA ORG
AMA ORG
Last spring DATE
AMA ORG
COVID ORG
FDA ORG
the end of October DATE
AMA ORG
COVID ORG
last spring DATE
COVID ORG
the 1950s DATE
one CARDINAL
EKG ORG
African NORP
Trump PERSON
Trump ORG
White House ORG
Trump ORG
Fox News ORG
Neil Cavuto PERSON
VA ORG
Lancet ORG
New England Journal ORG
Avastin PERSON
15 years DATE
FDA ORG
Avastin PERSON
Ketamine PERSON
FDA ORG
1970 DATE
Neil Cavuto PERSON
187 CARDINAL
122 CARDINAL
100 percent PERCENT
COVID ORG
this year DATE
AMA ORG
the end of October DATE

AMA ORG
last summer DATE
COVID ORG
three CARDINAL
COVID ORG
AMA ORG
AMA ORG
Oct. 30 DATE
just a few days DATE
Democrats NORP
Trump ORG
COVID ORG
Trump ORG
AMA ORG
AMA ORG
Hunter Biden's PERSON
ICU ORG
AMA ORG
AMA ORG
Orange Man ORG
the White House FAC
The American Medical Association ORG
the past year DATE
Just a few days DATE
the past year DATE
AMA ORG
only 12 percent PERCENT
AMA ORG
AMA ORG
AMA ORG
Last spring DATE
AMA ORG
COVID ORG
FDA ORG
the end of October DATE
AMA ORG
COVID ORG
last spring DATE
COVID ORG
the 1950s DATE
one CARDINAL
EKG ORG
African NORP
Trump PERSON
Trump ORG
White House ORG
Trump ORG
Fox News ORG
Neil Cavuto PERSON
VA ORG

Lancet ORG
New England Journal ORG
Avastin PERSON
15 years DATE
FDA ORG
Avastin PERSON
Ketamine PERSON
FDA ORG
1970 DATE
Neil Cavuto PERSON
187 CARDINAL
122 CARDINAL
100 percent PERCENT
COVID ORG
this year DATE
AMA ORG
the end of October DATE
AMA ORG
last summer DATE
COVID ORG
three CARDINAL
COVID ORG
AMA ORG
AMA ORG
Oct. 30 DATE
just a few days DATE
Democrats NORP
Trump ORG
COVID ORG
Trump ORG
AMA ORG
AMA ORG
Hunter Biden's PERSON
ICU ORG
AMA ORG
AMA ORG
Orange Man ORG
the White House FAC
UN Supports Disinformation Campaign ORG
New State Budget Is New Yorkers' Worst Nightmare ORG
State ORG
Apollo ORG
Alien City GPE
1984 DATE
the New York Times ORG
Trump ORG
The World Health Organization ORG
months DATE
WHO ORG
WHO ORG

34 CARDINAL
Monday October 5th DATE
Michael Ryan PERSON
WHO ORG
roughly 10% PERCENT
around 35 million CARDINAL
Margaret Harris PERSON
WHO ORG
WHO ORG
Ryan PERSON
roughly 7.8 billion CARDINAL
10% PERCENT
780 million CARDINAL
1,061,539 CARDINAL
0.14% PERCENT
0.14% PERCENT
24 CARDINAL
WHO ORG
3.4% PERCENT
March DATE
Covid PERSON
IFR ORG
0.14% PERCENT
Covid PERSON
Dr Ryan's PERSON
WHO ORG
Covid NORP
Pfizer PERSON
Jay Leno PERSON
Republican NORP
Trump Supporter PERSON
71 DATE
Google ORG
Santiam Correctional Institution ORG
last month DATE
JAMA GPE
monthly DATE
the American Medical Association ORG
COVID-19 ORG
15% PERCENT
John Aitken PERSON
San Antonio Fox 8 GPE
COVID-19 ORG
COVID-19 ORG
Florida GPE
Nicky Morgan PERSON
Flynn PERSON
Netflix Clarifies Its Content Is 'Mostly Pedophilia-Free' ORG
Kremlin ORG
German Clinic's ORG

Diagnosis of Navalny's Condition ORG
Netizens ORG
ngayon ORG
China GPE
NIPOST ORG
Osogbo GPE
Wear a Mask WORK_OF_ART
Adam Hasner PERSON
RESPONDS ORG
London GPE
May DATE
City Council ORG
the U.S. Centers for Disease Control and Prevention ORG
CDC ORG
the Vaccine Adverse Event Reporting System ORG
more than 10 years DATE
at least 108 CARDINAL
today DATE
today DATE
first ORDINAL
U.S. GPE
2003 DATE
the Associate Press ORG
Anne Schuchat PERSON
CDC ORG
National Center for Immunization and Respiratory Diseases ORG
VAERS ORG
at least 108 CARDINAL
2003 DATE
96 CARDINAL
MMR ORG
first ORDINAL
1963 DATE
U.S. GPE
the years DATE
first ORDINAL
1963 DATE
the United States GPE
approximately 98% PERCENT
the International Medical Council on Vaccination ORG
first ORDINAL
IMCV PERSON
Fani-Kayode PERSON
Abba Kyari's PERSON
Paul Ryan PERSON
Seven CARDINAL
Radio Ad Campaign against Massive Immigration Increases ORG
Amnesty ORG
The New York Times ORG
The World Health Organization ORG

months DATE
WHO ORG
WHO ORG
34 CARDINAL
Monday October 5th DATE
Michael Ryan PERSON
WHO ORG
roughly 10% PERCENT
around 35 million CARDINAL
Margaret Harris PERSON
WHO ORG
WHO ORG
Ryan PERSON
roughly 7.8 billion CARDINAL
10% PERCENT
780 million CARDINAL
1,061,539 CARDINAL
0.14% PERCENT
0.14% PERCENT
24 CARDINAL
WHO ORG
3.4% PERCENT
March DATE
Covid PERSON
IFR ORG
0.14% PERCENT
Covid PERSON
Dr Ryan's PERSON
WHO ORG
Covid NORP
DEMOCRATS NORP
Michigan GPE
Biden PERSON
an estimated 6,000 CARDINAL
Trump ORG
Oregon Legislature ORG
more than 1,600 CARDINAL
Feb. 1 DATE
Biden Accepts Campaign Donation ORG
Greenland GPE
London GPE
Greenwich GPE
Lambeth PERSON
The Phony War Against CO2
 ORG
UN Supports Disinformation Campaign ORG
18-Year-Old DATE
U.S. GPE
CBP ORG
ICE Custody for ORG

Three Weeks DATE
U.S. GPE
NAMES ORG
ON CHINA COMMUNIST PARTY ORG
1st ORDINAL
WTO ORG
22 CARDINAL
more than 55 CARDINAL
Coroner Saying George Floyd Died of Drug Overdose ORG
The Washington Sentinel GPE
Doug Whitsett PERSON
Oregon GPE
Joy Covey PERSON
Amazon ORG
50 CARDINAL
Woodside GPE
British NORP
100 years ago DATE
Anthony Fauci PERSON
Trump PERSON
2005 DATE
the National Institutes of Health ORG
NIH ORG
COVID-19 ORG
79% PERCENT
The Virology Journal ORG
Fauci's PERSON
National Institutes of Health ORG
August 22, 2005 DATE
Emphasis ORG
Anthony PERSON
Fauci PERSON
15 years DATE
Said PERSON
Fauci PERSON
NIH ORG
2005 DATE
10 CARDINAL
Fauci PERSON
Didier Raoult PERSON
the Anthony Fauci ORG
France GPE
February 25 DATE
79 CARDINAL
80 CARDINAL
COVID-19 ORG
VA ORG
first ORDINAL
Andrew Cuomo PERSON
New York GPE

March 6 DATE
Democrat NORP
Nevada GPE
Michigan GPE
March 28 DATE
U.S. GPE
March 20 DATE
Trump PERSON
FDA ORG
March 23 DATE
Vladimir Zelenko PERSON
500 CARDINAL
100% PERCENT
Fauci PERSON
last Thursday DATE
approximately 350 CARDINAL
Kiryas Joel ORG
150 CARDINAL
New York GPE
ZERO ORG
ZERO ORG
ZERO ORG
approximately 10% PERCENT
Said PERSON
Zelenko PERSON
about $20 MONEY
20,000 MONEY
Fauci PERSON
2005 DATE
Zelenko PERSON
first ORDINAL
Jesus PERSON
first ORDINAL
one CARDINAL
Fauci PERSON
April 9 DATE
33,000 CARDINAL
1934 DATE
Jonas Salk PERSON
1954 DATE
Fauci PERSON
Raoult PERSON
Zelensky PERSON
20 MONEY
Fauci PERSON
the Talking Snake Media ORG
Delphi ORG
15 minutes TIME
zero CARDINAL
payday DATE

Fauci ORG
February 15 DATE
Fauci PERSON
Coronavirus GPE
multi-billion dollar MONEY
day one DATE
today DATE
GEORGIA GPE
Arctic LOC
Matter Website PERSON
Biden PERSON
Son PERSON
Daughter PERSON
Vos PERSON
Fitzgerald PERSON
Evers PERSON
Wisconsin GPE
NHS ORG
Europe LOC
Arctic LOC
2017 DATE
National Insurance ORG
Obama's Daughters Caught ORG
Wisconsin Election Commission ORG
Ballots PERSON
Warren Statement PERSON
Boeing ORG
today DATE
James I about PERSON
four centuries ago DATE
Duane Carr PERSON
the University of Tennessee College of Medicine ORG
Victor Buhler PERSON
St. Joseph Hospital FAC
Kansas City GPE
thousands CARDINAL
Sheldon Sommers PERSON
Pathologist NORP
Lenox Hill Hospital ORG
New York GPE
Brigitte ORG
Youtube PERSON
German NORP
Rich White's PERSON
Smoke Screens PERSON
Jan Zeldenrust PERSON
Dutch NORP
the Government of Holland ORG
1951 — 1984 DATE
1980 DATE

Dutch NORP
these days DATE
Coal ORG
First ORDINAL
emphysema PERSON
4 CARDINAL
5 CARDINAL
4 CARDINAL
8) CARDINAL
first ORDINAL
emphysema PERSON
secondly ORDINAL
emphysema PERSON
Delhi GPE
'Delhi GPE
Delhi GPE
Indian NORP
Delhi GPE
Indian NORP
Delhi GPE
'Delhi GPE
emphysema PERSON
one CARDINAL
BioFlex ORG
firstly ORDINAL
secondly ORDINAL
the University of Arkansas ORG
Villanova University ORG
Marquette University ORG
more than 500 CARDINAL
U.S. GPE
Canadian NORP
Scot Burton PERSON
the Sam M. Walton College of Business ORG
two CARDINAL
Dem NORP
Knope Caught PERSON
Ballots GPE
Kimchi May Have Helped Lower PERSON
COVID-19 Fatalities ORG
European NORP
America GPE
Trump PERSON
last week DATE
Americans NORP
Medicare ORG
Social Security ORG
two CARDINAL
Trump PERSON
Republican NORP

Trump PERSON
Trump PERSON
21st Century DATE
2008 DATE
Congress ORG
some $700 billion MONEY
Treasury ORG
trillions CARDINAL
the Federal Reserve ORG
Ilhan Omar Tells PERSON
Minnesota LOC
Wisconsin GPE
Miraculous Vanishing Ballot Rejection Rate ORG
Two-Thirds CARDINAL
UK GPE
under 18 DATE
Apollo ORG
Alien City GPE
CDC Inflated COVID Numbers ORG
NASA ORG
Earth LOC
Alex Trebek Becomes PERSON
U.S. GPE
Abe Lincoln PERSON
Arrested Suspect Is A Black Lives Matter ORG
New Mexico House Committee ORG
Legalize Killing Babies ORG
Birth GPE
Pittsburg Area Antifa LOC
Tweets Warning PERSON
Trump WORK_OF_ART
Sunday DATE
noon TIME
Britain GPE
last seven years DATE
one CARDINAL
50 CARDINAL
Rainsy PERSON
163 CARDINAL
George Soros For PERSON
4 CARDINAL
National Review ORG
Uber PERSON
New Yorkers NORP
2005 DATE
Joe Biden Tells PERSON
Joe Biden PERSON
U.K. GPE
Obama GPE
Wisconsin GPE

Miraculous Vanishing Ballot Rejection Rate ORG
1000s MONEY
Accuse Roy Moore PERSON
Khrushchev NORP
America GPE
1.8billion MONEY
US GPE
U.S. GPE
Iran GPE
Americans NORP
One CARDINAL
four CARDINAL
David Cameron PERSON
ISIS ORG
Syria GPE
Apocalyptic Sea-Level Rise ORG
Congressional Budget Office ORG
Eugenie PERSON
ninth ORDINAL
9.2bn MONEY
NHS ORG
Hancock ORG
the White House Arresting Congress FAC
China GPE
Canada GPE
Nicky Morgan PERSON
Ohio Senate ORG
Earth LOC
Rastafarian Prez PERSON
Nana Akufo-Addo ORG
Georgia GPE
GOP Areas ORG
DEM Areas PERSON
CDC ORG
COVID-19 ORG
U.S. GPE
CDC ORG
between March and December 2020 DATE
1.5 percent PERCENT
two months DATE
0.5 percent PERCENT
the first 20 days DATE
less than 2 percent PERCENT
100 days DATE
CDC ORG
Caruthers California: Video LOC
DAILY GPE
Mackowiak PERSON
Abbott PERSON
Austin PERSON

Twins ORG
Democrat NORP
NHS ORG
1bn ORDINAL
5 CARDINAL
5 CARDINAL
NYC ORG
ICU ORG
Cameron Kyle-Sidell PERSON
tens of thousands CARDINAL
New Yorkers NORP
30,000 feet QUANTITY
up to half a million CARDINAL
BLM Petition to Change the Name of ORG
the White House ORG
David Davis PERSON
Canada GPE
Tusk PERSON
Scottish Parliament ORG
Tories NORP
L.A. Times Op-ed ORG
California Bear Flag ORG
Boris Johnson PERSON
SNP ORG
Biden Accepts Campaign Donation ORG
New Research Study Clarifies Health Outcomes ORG
Will Austin PERSON
One CARDINAL
five CARDINAL
Sarah Parker PERSON
Supreme Court ORG
Bill Withers PERSON
$2 Million MONEY
Trump Re-election Campaign ORG
Harambe Received PERSON
The Washington Post ORG
Vladimir Putin PERSON
The Revolution Begins at Home EVENT
Only 3% PERCENT
Trump Voters ORG
Joe Biden PERSON
Clinton Foundation ORG
Kenneth McCormick PERSON
Vladimir Putin PERSON
second ORDINAL
COVID ORG
Toronto GPE
Today DATE
Pfizer Covid PERSON
More than one CARDINAL

10 CARDINAL
Broward County Sheriff Fired For Lying About Parkland ORG
Heathrow FAC
third ORDINAL
UK GPE
Rubio, Risch Statement On Senate Intelligence Committee Release Of
Interrogation Study ORG
CANCER IS CANDIDA/FUNGUS AND CAN BE CURED ORG
Andy Murray PERSON
John Inverdale PERSON
BBC ORG
Pennsylvania GPE
372,000 CARDINAL
72 Hours TIME
250,000 CARDINAL
Tory NORP
Biden PERSON
$3 Trillion MONEY
Plane GPE
Article 50 LAW
Britain GPE
EU ORG
AAA ORG
Oregon GPE
millions CARDINAL
1 CARDINAL
Earth LOC
400th FAC
One CARDINAL
four CARDINAL
Jeffrey Epstein PERSON
Existence PERSON
More Than 10,000 CARDINAL
Cast Ballots PERSON
Michigan GPE
Iran GPE
US GPE
London GPE
May DATE
Klobuchar PERSON
Hennepin County Attorney's GPE
George Floyd Death PERSON
Hunter Biden Laptop PERSON
Joe Biden PERSON
International Red Cross Report ORG
Jews NORP
Wasserman Schultz PERSON
2016 DATE
Senate ORG
650 CARDINAL

UK GPE
BRAZIL GPE
AstraZeneca ORG
Oxford University ORG
Anvisa ORG
one CARDINAL
three CARDINAL
Brazil GPE
Oxford University ORG
the UK Government ORG
100 million CARDINAL
US GPE
September 6 DATE
India GPE
South Africa GPE
AstraZeneca ORG
Europe LOC
Inclusive Vaccine Alliance ORG
400 million CARDINAL
European NORP
UK GPE
30 million CARDINAL
BioNtech/Pfizer ORG
60 million CARDINAL
Valneva - in PERSON
six CARDINAL
UK GPE
Patrick Vallance PERSON
next spring DATE
the Commons and Lords National Security Strategy Committee ORG
AstraZeneca ORG
the turn of the year DATE
UK GPE
Vaccine Taskforce FAC
Kate Bingham PERSON
UK GPE
four CARDINAL
six CARDINAL
three CARDINAL
one CARDINAL
two CARDINAL
Covid PERSON
26,688 CARDINAL
UK GPE
789,229 CARDINAL
191 CARDINAL
today DATE
Democrat NORP
The Washington Post ORG
Chelsea Clinton PERSON

18th ORDINAL
Pervs ORG
five CARDINAL
School ORG
four years DATE
five CARDINAL
Greg Abbott PERSON
Texas GPE
billions CARDINAL
Democrats NORP
Social Security and Military ORG
Hank Aaron PERSON
Moderna COVID GPE
Jan. 5 DATE
the New York Times ORG
Atlanta Braves ORG
86-year-old DATE
Hall of Famer's ORG
today DATE
CNN ORG
Aaron PERSON
Aaron PERSON
earlier this month DATE
Moderna PERSON
the Associated Press ORG
Aaron PERSON
the Morehouse School of Medicine ORG
Atlanta GPE
Black Americans NORP
AP ORG
first ORDINAL
two CARDINAL
the Morehouse School of Medicine ORG
African Americans NORP
Tuskegee ORG
U.S. GPE
Black LOC
December DATE
VOX ORG
COVID ORG
Germany GPE
Norway GPE
43 CARDINAL
Pfizer PERSON
China GPE
Aaron PERSON
Robert F. Kennedy PERSON
Jr. PERSON
Children's Health Defense ORG
Aaron PERSON

COVID ORG
Kennedy PERSON
the Vaccine Adverse Events Reporting System ORG
the U.S. Department of Health and Human Services ORG
HHS ORG
2001 DATE
HHS ORG
fewer than 1% PERCENT
last week DATE
California GPE
330,000 CARDINAL
Moderna PERSON
Defender PERSON
56-year-old DATE
Florida GPE
about two weeks DATE
first ORDINAL
Pfizer ORG
Florida GPE
Pfizer ORG
32.2bn MONEY
just ONE year DATE
a quarter DATE
just five years DATE
Uber PERSON
Those Non-Disappearing Pacific Islands ORG
Portland GPE
Unknowns GPE
Egypt GPE
The New York Times ORG
Disneyland FAC
California GPE
Texas GPE
Scottish Parliament ORG
Tories NORP
UK GPE
Florida GPE
Rick Scott PERSON
Paris Climate Agreement ORG
Americans NORP
Feds Handcuff 'Irrational' House PERSON
Raiding Her Office ORG
Scott Walker's PERSON
Pennsylvania GPE
372,000 CARDINAL
7,000 CARDINAL
Mueller ORG
RT America's ORG
Sanchez GPE
Manuka Honey Kills PERSON

Arctic LOC
Schumer PERSON
Biden PERSON
50 MONEY
125 MONEY
Biden Has Been PERSON
Jerome Corsi PERSON
Wisconsin GPE
U.S. GPE
Baldwin PERSON
Las Vegas GPE
up to half a million CARDINAL
KRAKEN ORG
CIA Hacking Program ORG
United States GPE
Among Least Polluting Nations ORG
the Planet LOC
2017 DATE
US GPE
Miles PERSON
George Soros PERSON
Zakharova ORG
New Salisbury Claims Aim to Distract Public From the Incident ORG
GEORGIA GPE
132,000 CARDINAL
Ballots ORG
Fulton County GPE
Georgia GPE
Devon PERSON
One CARDINAL
four CARDINAL
Tom Hanks PERSON
65 DATE
Passes Away PERSON
Olympics EVENT
Rio Games ORG
Trans-Pecos Pipeline ORG
Presidio Crossing Pmject PERSON
Scott Walker PERSON
Mary Burke PERSON
Buy Isa ORG
500,000 CARDINAL
first ORDINAL
Arctic LOC
Mexico GPE
Barack Obama PERSON
75 CARDINAL
MPH ORG
This Winter DATE
Hise PERSON

Ghana GPE
2nd ORDINAL
Africa - Report LOC
Vladimir Putin PERSON
Europe LOC
Disneyland FAC
California GPE
Texas GPE
Paul Nuttall PERSON
Stoke LOC
99.96% PERCENT
two CARDINAL
Pfizer ORG
Covid PERSON
Israel GPE
Cancer Discovery PERSON
two CARDINAL
Leopoldo Segal PERSON
the Lung Microbiome Program ORG
New York University Grossman School of Medicine ORG
IL-17 PRODUCT
IL-17 PRODUCT
Segal PERSON
83 CARDINAL
Veillonella GPE
Prevotella ORG
Streptococcus ORG
IL-17 PRODUCT
One CARDINAL
Canadian NORP
Aji Joffe PERSON
at least ten CARDINAL
Harvard ORG
Duke ORG
John Hopkins Universities PERSON
COVID-19 ORG
0.89 CARDINAL
1.37 million CARDINAL
the next 15 and 20 years DATE
as high as 50% PERCENT
just a few months DATE
June DATE
Earth LOC
European NORP
100-Year-Old DATE
Capitol ORG
Capitol Shooting Exposed ORG
Hollywood GPE
The Squad Wants Bible Deemed To Be 'Hate Speech ORG
Coroner Saying George Floyd Died of Drug Overdose ORG

The Washington Sentinel GPE
Ministry of Ayurveda, Yoga & Naturopathy ORG
Unani PERSON
Siddha PERSON
Homoeopathy PERSON
Ayurveda ORG
Siddha PERSON
Ross McKitrick PERSON
America GPE
PCR ORG
the past few months DATE
PCR ORG
COVID-19 ORG
COVID-19 ORG
JustTheNews NORP
October DATE
COVID-19 ORG
PCR ORG
COVID-19 ORG
PJMedia ORG
Stacey Lennox ORG
Biden PERSON
CDC ORG
FDA ORG
Americans NORP
Uncle Joe PERSON
months DATE
TDS ORG
Lennox ORG
100% PERCENT
as less than one hour DATE
Biden PERSON
WHO ORG
August of last year DATE
The New York Times ORG
as many as 90% PERCENT
COVID-19 ORG
three CARDINAL
PCR ORG
the United States GPE
37-40 CARDINAL
around 30 CARDINAL
CDC ORG
COVID-19 ORG
PCR ORG
Times ORG
CDC ORG
PCR ORG
33 CARDINAL
clinicians NORP

PCR ORG
CDC ORG
five CARDINAL
July 13, 2020 DATE
40 DATE
September 28, 2020 DATE
Clinical Infectious Diseases ORG
Jaafar GPE
al. GPE
nearly 4,000 CARDINAL
30 CARDINAL
CDC ORG
PCR ORG
December 1, 2020 DATE
40 CARDINAL
the New York Times ORG
CDC ORG
COVID-19 ORG
Fauci PERSON
the Mendacious Midget ORG
August DATE
one CARDINAL
Fauci PERSON
late January DATE
one CARDINAL
Fauci ORG
Today DATE
an hour TIME
Joe Biden PERSON
WHO ORG
PCR ORG
Diagnostic LOC
1 CARDINAL
NAT ORG
one hour TIME
Biden PERSON
WHO ORG
PCR ORG
Andy Swan PERSON
January 20, 2021 DATE
Fauci PERSON
July DATE
35 CARDINAL
35 DATE
Jaafar NORP
Florida GPE
December DATE
WHO ORG
PCR ORG
PCR ORG

higher than 30 CARDINAL
the New York Times ORG
Jaafar NORP
CNN ORG
Times ORG
August DATE
approximately 2.43 million CARDINAL
24.3 million CARDINAL
COVID-19 ORG
September DATE
Joe Biden PERSON
Anthony Fauci PERSON
a year DATE
Andy Swan PERSON
January 20, 2021 DATE
this season DATE
COVID ORG
European NORP
Trump ORG
second Impeachment Trial ORG
the day DATE
House ORG
U.S. GPE
Utah GPE
3rd ORDINAL
Jason Chaffetz PERSON
the House Rules ORG
OAN ORG
Jack Posobiec PERSON
Donald Trump PERSON
second ORDINAL
Senate ORG
January 6 DATE
Congress ORG
Capitol Shooting Exposed ORG
Hollywood GPE
Washington State University ORG
100 CARDINAL
two CARDINAL
the Journal of Antimicrobial Chemotherapy ORG
Applied and Environmental Microbiology ORG
C. NORP
Mediterranean LOC
Garlic PERSON
Xiaonan Lu PERSON
One CARDINAL
Allium PRODUCT
Louis Pasteur PERSON
1858 DATE
the Middle Ages LOC

Russians NORP
Russian NORP
Bloomberg PERSON
The Fight for Wisconsin's Soul - The New York Times ORG
Rubio Hails PERSON
Bill To Defund PERSON
Vladimir Putin PERSON
Europe LOC
Jeffrey Epstein PERSON
Tesco ORG
119 CARDINAL
last year DATE
Flu Vaccine Deaths PERSON
South Korea GPE
nearly 100 CARDINAL
EU ORG
NY Assembly ORG
Democrats NORP
Gold Star ORG
Millions CARDINAL
the Atlantic Ocean LOC
Shifts Hint PERSON
Biden Has Been PERSON
Jerome Corsi PERSON
Democrats NORP
Bill To 'Euthanize Seniors' PERSON
Santiam Correctional Institution ORG
Texas GPE
Pervs ORG
five CARDINAL
School ORG
four years DATE
five CARDINAL
UK GPE
El Nino ORG
Domino PERSON
Earth LOC
Paul Stekler PERSON
Morgue GPE
Mueller ORG
Russia GPE
more than $25 million MONEY
The Fight for Wisconsin's Soul - The New York Times ORG
Trump Leading PERSON
Antarctica LOC
billions CARDINAL
One CARDINAL
three CARDINAL
Democratic NORP
Southern States GPE

China GPE
October DATE
Britain GPE
100 MONEY
Syria GPE
Cameron ORG
1bn ORDINAL
2055 CARDINAL
Tony Abbott PERSON
NBA ORG
Trump: Resign or We'll Never Play Again@usa news ORG
The Novel Coronavirus Infecting ORG
China GPE
Army ORG
Khrushchev NORP
America GPE
the COVID-19 Relief Fund ORG
Democrats NORP
ObamaCare ORG
Carey PERSON
15 CARDINAL
Nigerian House of Reps ORG
Coronavirus PERSON
Gossip PERSON
Nigeria GPE
TEN CARDINAL
TIMES ORG
Earth LOC
NASA ORG
One CARDINAL
Tetyana Obukhanych ORG
Ph.D. WORK_OF_ART
the Rockefeller University ORG
New York GPE
Harvard ORG
British Columbia GPE
Quebec GPE
2011 DATE
48% PERCENT
PhD WORK_OF_ART
one CARDINAL
two CARDINAL
a few years ago DATE
American NORP
first ORDINAL
two CARDINAL
11 DATE
Ukraine GPE
PhD WORK_OF_ART
Iran GPE

Erbil ORG
Iraq GPE
Senate ORG
Republican NORP
Democrats NORP
NY Post's 'Smoking Gun ORG
Hunter Biden PERSON
Forensic Analysis Concludes ORG
Mackowiak PERSON
Abbott PERSON
Austin PERSON
Government ORG
Washington State GPE
9 CARDINAL
Wolf PERSON
Britain GPE
six weeks DATE
Coronavirus PERSON
Joe Biden PERSON
Florida GPE
Rising Six Times ORG
2017 DATE
England GPE
first ORDINAL
Austin PERSON
USC ORG
Joel Hay PERSON
George Clooney Smack PERSON
the Middle of Epstein's LOC
Sex Cabal WORK_OF_ART
A decade DATE
Scottish National Party ORG
Joy Covey PERSON
Amazon ORG
50 CARDINAL
Woodside GPE
Austin GPE
Monday morning TIME
Pfizer ORG
90 percent PERCENT
Trump ORG
Operation Warp Speed ORG
the New York Times ORG
Washington Post ORG
Trump ORG
Times ORG
Kathrin Jansen PERSON
Pfizer ORG
US GPE
Pfizer PERSON

Operation Warp Speed ORG
NPR ORG
July DATE
nearly two billion dollars MONEY
Pfizer ORG
July DATE
U.S. GPE
100 million CARDINAL
$1.95 billion MONEY
Pfizer ORG
Jansen PERSON
Newsweek ORG
Jansen PERSON
Pfizer NORP
Operation Warp Speed ORG
First ORDINAL
Donald Trump PERSON
ten-second TIME
Google ORG
Pfizer PERSON
Operation Warp Speed ORG
Wikipedia GPE
Pfizer PERSON
Operation Warp Speed ORG
thousands CARDINAL
42nd street FAC
the New York Times ORG
Donald Trump PERSON
millions CARDINAL
less than two hours TIME
Operation Warp Speed ORG
Donald Trump PERSON
The Prices Of Food Items Are Coming Down ORG
Garba Shehu PERSON
Vitamin ORG
The Washington Post ORG
half CARDINAL
USCIS ORG
Boko Haram PERSON
Shekau PERSON
9 CARDINAL
Wis. GPE
Mich. GPE
Shanghai Government Officially Recommends Vitamin C ORG
COVID-19 ORG
25,000 CARDINAL
2020 DATE
Biden Introduces Granddaughter ORG
Beau Biden PERSON
Rubio Comments ORG

Three-quarters CARDINAL
NHS ORG
overnight TIME
Barack Obama PERSON
Rebecca Long-Bailey PERSON
Labour ORG
Ross McKitrick PERSON
Nolte PERSON
TRUMP ORG
Obama GPE
Reagan PERSON
Same Point ORG
Presidencies FAC
New Eavesdropping Equipment Sucks All Data Off Your Phone ORG
£30 billion MONEY
NHS ORG
RFK ORG
The Science" Video and Other Facts • Children's Health Defense
WORK_OF_ART
The New York Times ORG
CLARKSBURG ORG
Jared Henry ORG
Maddie Mansell PERSON
Buckhannon Pike PERSON
Nutter Fort GPE
Mansell PERSON
Henry PERSON
Mansell PERSON
Mansell PERSON
Nikki Bowmar PERSON
the U.S. Small Business Administration ORG
Bowmar PERSON
approximately 114,391 CARDINAL
West Virginia GPE
99 percent PERCENT
280,213 CARDINAL
nearly 50 percent PERCENT
Bowmar PERSON
Mansell PERSON
first ORDINAL
Christ GPE
George "Gib" Brown PERSON
The James and Law Co. ORG
West Virginia GPE
nearly 30 CARDINAL
Clarksburg GPE
two CARDINAL
1903 DATE
Brown PERSON
James and Law ORG

Brown PERSON
the years DATE
Chamber ORG
the Robinson Grand Performing Arts Center ORG
Clarksburg GPE
Clarksburg GPE
years DATE
Brown PERSON
Dan Taylor PERSON
the West Virginia Community Development Hub's Energizing
Entrepreneurial Communities ORG
the last of three years DATE
four CARDINAL
Grafton ORG
Taylor PERSON
Taylor PERSON
Stephen Mann PERSON
Lewis County GPE
December DATE
Weston GPE
Mann ORG
two CARDINAL
one CARDINAL
401k PRODUCT
Mann PERSON
Manns PERSON
Lewis County GPE
Mann PERSON
Lewis County GPE
West Virginia GPE
Bowmar PERSON
Mann ORG
The Energizing Entrepreneurial Communities ORG
Grafton ORG
Taylor PERSON
All Aboard Grafton FAC
Taylor PERSON
Taylor County PERSON
first ORDINAL
February DATE
Patricia Henderson PERSON
the Taylor County Development Authority ORG
the Taylor County Commission ORG
Taylor County GPE
Weston GPE
Fairs and Festivals Committee ORG
Julia Spelsberg PERSON
Spelsberg PERSON
Weston GPE
North Central West Virginia LOC

The Small Business Administration ORG
Bowmar PERSON
USDA ORG
the Tax Cut FAC
Taylor PERSON
NASA ORG
California GPE
the Changing Climate - The New York Times ORG
Breonna Taylor PERSON
SHE ORG
Georgia Voters Sound ORG
The Alarming Thing About Climate Alarmism ORG
Charlie Hebdo's PERSON
West LOC
Islam NORP
Noa Pothoven PERSON
IOC ORG
Beijing GPE
2022 DATE
over 30% PERCENT
Murders GPE
States GPE
Up to a third CARDINAL
David McKinley PERSON
2017 DATE
Philip Hammond PERSON
Tory NORP
National Insurance ORG
1bn MONEY
next two years DATE
Europe LOC
ObamaCare ORG
Carey PERSON
Erpenbach PERSON
Walker PERSON
Health Insurance Exchanges ORG
Florida GPE
Rick Scott PERSON
Paris Climate Agreement ORG
Americans NORP
Julia Roberts: 'Michelle Obama PERSON
Toilet ORG
Hise PERSON
NHS ORG
Oregon Legislature ORG
more than 1,600 CARDINAL
Feb. 1 DATE
Senate ORG
Republican NORP
Ortt PERSON

Republican NORP
95-Degree Days DATE
Howard University Receives ORG
$4 Million MONEY
Hopper-Dean Foundation Gift ORG
Expand ORG
Bison STEM Scholars Program — HBCU Times ORG
GEORGIA GPE
132,000 CARDINAL
Ballots ORG
Fulton County GPE
Georgia GPE
Conor Lamb PERSON
15 CARDINAL
Nigerian House of Reps ORG
Coronavirus PERSON
Gossip PERSON
Nigeria GPE
Government ORG
Ross McKitrick PERSON
Clinton Foundation ORG
22,000 CARDINAL
U.S. GPE
Americans NORP
3 CARDINAL
China GPE
Myanmar GPE
Vitamin C ORG
Vermont GPE
Narcan ORG
FG ORG
England GPE
Wales GPE
Justine Greening PERSON
Lancashire GPE
17th DATE
9 days DATE
9.2bn MONEY
NHS ORG
Hancock ORG
NHS ORG
15,000 CARDINAL
six years DATE
Exclusive: The TCF Center Election Fraud — Newly Discovered Video
Shows Late Night Deliveries WORK_OF_ART
Tens of Thousands CARDINAL
8 Hours TIME
L.A. Times Op-ed ORG
California Bear Flag ORG
ICT ORG

Biden PERSON
Russia GPE
SolarWinds ORG
98% PERCENT
Rallying the World WORK_OF_ART
Defeat Malaria FAC
Normandy Four PERSON
Doug Whitsett PERSON
Oregon GPE
U.S. GPE
Jeff Merkley PERSON
Oregon GPE
Next year DATE
the year DATE
Arctic LOC
UK GPE
Central Park's FAC
COVID-19 FAC
Barack Obama PERSON
Boko Haram PERSON
Shekau PERSON
UK GPE
under 18 DATE
Warren Statement PERSON
Boeing ORG
52% PERCENT
Americans NORP
U.S. GPE
Jeff Merkley PERSON
Oregon GPE
U.S. GPE
Tammy Baldwin PERSON
Wisconsin GPE
NHS ORG
two CARDINAL
each DAY DATE
800,000 years DATE
Turkey GPE
Pakistan GPE
India GPE
NHS ORG
seven years DATE
Fauci Hold PERSON
ICT ORG
Court ORG
Ameen Jogee ORG
Fisher-Price Releases 'My First Peaceful Protest' Playset With House
ORG
Justine Greening PERSON
ACLU ORG

Florida Responds ORG
U.S. Supreme Court ORG
El Niño ORG
2015 DATE
Fani-Kayode PERSON
Abba Kyari's PERSON
over 4 million CARDINAL
the United States GPE
Rick Simpson PERSON
marijuana PERSON
Cannabis PERSON
Marijuana PERSON
7 CARDINAL
Atlanta GPE
Judicial Watch ORG
Detail Role PERSON
Justice Department ORG
Trayvon Martin PERSON
Travis Rowley PERSON
Democrats NORP
first ORDINAL
78 Year Old DATE
CIA ORG
Mary Burke PERSON
5,000 CARDINAL
One CARDINAL
10 CARDINAL
Austin GPE
Donald Trump's PERSON
Rainsy PERSON
Kwik PERSON
Launches Kwik Strip Gentlemen's PERSON
CDC ORG
COVID-19 ORG
U.S. GPE
CDC ORG
between March and December 2020 DATE
1.5 percent PERCENT
two months DATE
0.5 percent PERCENT
the first 20 days DATE
less than 2 percent PERCENT
100 days DATE
CDC ORG
Tom Hanks PERSON
65 DATE
Passes Away PERSON
Coronavirus Contains HIV Insertions ORG
Jeremy Corbyn PERSON
150,000 CARDINAL

Donald Trump's PERSON
Britain GPE
12,000 CARDINAL
Chris King Statement PERSON
800 Percent PERCENT
Health Insurance Premiums for Low-Income Seniors ORG
Trumpcare ORG
Orlando GPE
Florida GPE
Trump ORG
Earth LOC
Didier Raoult PERSON
1,061 CARDINAL
91% PERCENT
Raoult PERSON
Hydroxychloroquine Azithromycin PERSON
COVID-19 ORG
0.5% PERCENT
first ORDINAL
Didier Raoult PERSON
French NORP
100% PERCENT
Hydroxychloroquine PERSON
Azithromycin PERSON
6 days DATE
80 CARDINAL
one 86 year-old DATE
74-year old DATE
one CARDINAL
the age of 86 DATE
today DATE
IHU Méditerranée Infection FAC
Tonight TIME
Raoult PERSON
OZ PERSON
Didier Raoult PERSON
1000 CARDINAL
99.3% PERCENT
1000 CARDINAL
20 CARDINAL
ICU ORG
7 CARDINAL
Raoult PERSON
Raoul PERSON
21:42 minutes TIME
one CARDINAL
Mexico GPE
European Court of Killers' Rights ORG
New Eavesdropping Equipment Sucks All Data Off Your Phone ORG
RFK ORG

The Science" Video and Other Facts • Children's Health Defense
WORK_OF_ART
New York's GPE
2018 DATE
GEORGIA GPE
132,000 CARDINAL
Ballots ORG
Fulton County GPE
Georgia GPE
1000s MONEY
Accuse Roy Moore PERSON
Warren Statement PERSON
Boeing ORG
Vermont GPE
Narcan ORG
62% PERCENT
EFCC ORG
Theodore Orji PERSON
N521 billion MONEY
FDA Shocking Study: Cells Used ORG
Biden Introduces Granddaughter ORG
Beau Biden PERSON
Schools ORG
1.3bn MONEY
United States GPE
Among Least Polluting Nations ORG
the Planet LOC
Kremlin ORG
German Clinic's ORG
Diagnosis of Navalny's Condition ORG
72 Hours TIME
Food Stamp Enrollment Drops ORG
Four Million CARDINAL
One Month DATE
Fox News ORG
Jeanine Show PERSON
Those Non-Disappearing Pacific Islands ORG
Domino PERSON
Earth LOC
George Osborne's PERSON
2015 DATE
Trump ORG
COVID-19 ORG
NASA ORG
Earth LOC
Thousands CARDINAL
BANK BUSTING Labour's ORG
four CARDINAL
10billion MONEY
Green Party ORG

Natalie Bennett PERSON
Ofcom ORG
Kremlin ORG
U.S. GPE
Wisconsin GPE
Miraculous Vanishing Ballot Rejection Rate ORG
Minneapolis GPE
George Floyd PERSON
Derek Chauvin PERSON
Trump ORG
Earth LOC
Capitol Shooting Exposed ORG
Hollywood GPE
US GPE
ISIS ORG
Syrian NORP
Labour ORG
100,000 CARDINAL
3 years DATE
Only 3% PERCENT
Trump Voters ORG
Joe Biden PERSON
TEN CARDINAL
TIMES ORG
Earth LOC
NASA ORG
£570 billion MONEY
the NHS Long Term Plan LOC
CO2 CARDINAL
Putin PERSON
32.2bn MONEY
just ONE year DATE
a quarter DATE
just five years DATE
Trump Revokes All ORG
First ORDINAL
Capitol ORG
Six CARDINAL
Congress ORG
36% PERCENT
Candace Owens PERSON
NBA Players Wear Special Lace Collars To Honor Ruth Bader Ginsburg ORG
EU ORG
100,000 CARDINAL
New Yorkers NORP
Pittsburg Area Antifa LOC
Tweets Warning PERSON
Trump WORK_OF_ART
Sunday DATE
noon TIME

Windham GPE
New Hampshire GPE
300 CARDINAL
Votes NORP
Northern Ireland's GPE
Covid NORP
December DATE
the U.S. Food and Drug Administration ORG
FDA ORG
Emergency Use Authorization PERSON
two CARDINAL
RNA ORG
COVID-19 ORG
Americans NORP
Pfizer ORG
Moderna PERSON
the American College of Obstetricians ORG
Gynecologists ORG
Feb. 12 DATE
the Vaccine Adverse Event Reporting System ORG
111 CARDINAL
Pfizer PERSON
Moderna PERSON
first ORDINAL
Dec. 22 DATE
just 10 days DATE
Pfizer PERSON
Nearly a third CARDINAL
31% PERCENT
as little as one day CARDINAL
VAERS ORG
37-year-old DATE
first ORDINAL
Moderna PERSON
28 weeks DATE
just one week later DATE
35-year-old DATE
around 29 weeks DATE
two days DATE
Pfizer ORG
Two CARDINAL
first ORDINAL
one CARDINAL
Anthony Fauci PERSON
Jan. 21 DATE
Feb. 3 DATE
Fauci PERSON
January DATE
20,000 CARDINAL
10,000 CARDINAL

Fauci ORG
COVID ORG
only two months DATE
Biden PERSON
the National Institute of Allergy and Infectious Diseases ORG
Moderna PERSON
Moderna PERSON
The University of Washington ORG
COVID ORG
the World Health Organization ORG
Jan. 27 DATE
Moderna PERSON
two days later DATE
The New York Times ORG
Peter Breggin PERSON
seven-decade-old DATE
two decades DATE
Gates ORG
2009 DATE
2015-16 DATE
COVID-19 PERSON
Who's the Target WORK_OF_ART
the mid-2000s DATE
the Centers for Disease Control and Prevention ORG
CDC ORG
roughly three CARDINAL
five CARDINAL
American NORP
Tdap PERSON
two CARDINAL
five CARDINAL
Tdap PERSON
Tdap PERSON
one CARDINAL
Tdap PERSON
third ORDINAL
2016 DATE
the 21st Century DATE
Task Force on Research Specific ORG
Pregnant Women and Lactating Women ORG
the National Institutes of Health ORG
NIH ORG
COVID ORG
NIH ORG
Pfizer PERSON
Feb. 19 DATE
COVID ORG
4,000 CARDINAL
the European Medicines Agency ORG
EMA ORG

Dec. 1 DATE
Michael Yeadon PERSON
one CARDINAL
Pfizer ORG
Wolfgang Wodarg PERSON
German NORP
COVID ORG
a week DATE
December DATE
The Control Group ORG
1,248% PERCENT
1,099% PERCENT
Truvada GPE
China GPE
Russia GPE
2020 DATE
China GPE
Russia GPE
Hong Kong GPE
Germany GPE
Canada GPE
Czech Republic Hacked PA GPE
NV ORG
MI ORG
GA Battleground States Raw Data Analytics Show ORG
Mitt Romney PERSON
$1 million MONEY
Senate ORG
Utah GPE
Merck Scraps COVID Vaccines ORG
Australian NORP
Months DATE
the Texas House ORG
Texas Scorecard ORG
Empower Texans NORP
one CARDINAL
First Amendment LAW
January DATE
two CARDINAL
Texas Scorecard ORG
Brandon Waltens PERSON
Destin Sensky PERSON
Legislature ORG
Senate ORG
House ORG
two CARDINAL
House Administration Chair ORG
Charlie Geren PERSON
Fort Worth GPE
Republican NORP

Empower Texans NORP
PAC ORG
January DATE
Texas Scorecard ORG
PAC ORG
Geren PERSON
Geren PERSON
Empower Texans NORP
PAC ORG
Geren PERSON
Dan Patrick PERSON
the Texas Senate ORG
more than $850,000 MONEY
the last five years DATE
this week DATE
Austin GPE
Geren PERSON
Geren PERSON
two CARDINAL
the Texas House ORG
Wednesday DATE
morning TIME
Geren PERSON
Empower Texans NORP
weekday DATE
weekly DATE
A-F PRODUCT
the Texas House ORG
Waltens PERSON
Sensky PERSON
second ORDINAL
House ORG
Geren PERSON
Empower Texans NORP
2017 DATE
Geren PERSON
Cary Cheshire Waltens PERSON
Sensky ORG
the Texas Senate ORG
the Texas House ORG
OMINOUS ORG
U.S. GPE
Hal Turner PERSON
50,000 CARDINAL
Chinese NORP
Maine GPE
Devon PERSON
NHS ORG
thousands CARDINAL
George Osborne PERSON

Labour ORG
Commons ORG
3 million years DATE
98% PERCENT
Atlanta GPE
Scott Walker PERSON
$1 million MONEY
USA GPE
World Cup EVENT
Netherlands GPE
2 CARDINAL
Britain GPE
42 MONEY
Clinton Foundation ORG
Ilhan Omar Tells PERSON
Minnesota LOC
U.S. GPE
Americans NORP
3 CARDINAL
Wednesday 25 April 2018 DATE
Biden Quietly Pardons ORG
2,617 CARDINAL
051316 DATE
Emmett Till's Memorial Sign Shot PERSON
35 days DATE
Democrat Fundraising Machine 'ActBlue Texas ORG
Caught Paying Out Small Sums WORK_OF_ART
Hundreds of Individuals for Unknown Purposes - David Harris Jr ORG
Israeli NORP
first ORDINAL
year DATE
Dan Aridor PERSON
Accelerated Evolution Biotechnologies Ltd. ORG
2000 DATE
ITEK GPE
SoAP ORG
day one DATE
a few weeks DATE
Aridor PERSON
Ilan Moran PERSON
Ilan Moran PERSON
an estimated 18.1 million MONEY
each year DATE
the International Agency for Research on Cancer ORG
sixth ORDINAL
second ORDINAL
second ORDINAL
Aridor ORG
Ilan Morad PERSON
SoAP ORG

2018 DATE
the Nobel Prize WORK_OF_ART
two CARDINAL
Morad GPE
first ORDINAL
Morad GPE
Morad GPE
Hanan Itzhaki PERSON
Morad GPE
at least three CARDINAL
Morad GPE
Morad GPE
three CARDINAL
three CARDINAL
Morad GPE
Morad PERSON
Morad PERSON
12 CARDINAL
Morad PERSON
one CARDINAL
two CARDINAL
Morad GPE
Today DATE
two CARDINAL
Morad GPE
only a few weeks DATE
Aridor GPE
Morad PERSON
first ORDINAL
a few years DATE
Fauci Hold PERSON
Coronavirus ORG
Chinese NORP
NHS ORG
30% PERCENT
Merck Scraps COVID Vaccines ORG
Justine Greening PERSON
EU ORG
Trump Wins ORG
$500 Million MONEY
CNN ORG
2016 DATE
Presidential Election Volume I of II Special Counsel WORK_OF_ART
Robert S. Mueller PERSON
III Submitted Pursuant ORG
28 CARDINAL
Kirk Douglas Dead PERSON
4 Days DATE
Covid-19 ORG
Paris GPE

Al Gore PERSON
Monday morning TIME
Pfizer ORG
90 percent PERCENT
Trump ORG
Operation Warp Speed ORG
the New York Times ORG
Washington Post ORG
Trump ORG
Times ORG
Kathrin Jansen PERSON
Pfizer ORG
US GPE
Pfizer PERSON
Operation Warp Speed ORG
NPR ORG
July DATE
nearly two billion dollars MONEY
Pfizer ORG
July DATE
U.S. GPE
100 million CARDINAL
$1.95 billion MONEY
Pfizer ORG
Jansen PERSON
Newsweek ORG
Jansen PERSON
Pfizer NORP
Operation Warp Speed ORG
First ORDINAL
Donald Trump PERSON
ten-second TIME
Google ORG
Pfizer PERSON
Operation Warp Speed ORG
Wikipedia GPE
Pfizer PERSON
Operation Warp Speed ORG
thousands CARDINAL
42nd street FAC
the New York Times ORG
Donald Trump PERSON
millions CARDINAL
less than two hours TIME
Operation Warp Speed ORG
Donald Trump PERSON
EU ORG
100,000 CARDINAL
America GPE
PCR ORG

the past few months DATE
PCR ORG
COVID-19 ORG
COVID-19 ORG
JustTheNews NORP
October DATE
COVID-19 ORG
PCR ORG
COVID-19 ORG
PJMedia ORG
Stacey Lennox ORG
Biden PERSON
CDC ORG
FDA ORG
Americans NORP
Uncle Joe PERSON
months DATE
TDS ORG
Lennox ORG
100% PERCENT
as less than one hour DATE
Biden PERSON
WHO ORG
August of last year DATE
The New York Times ORG
as many as 90% PERCENT
COVID-19 ORG
three CARDINAL
PCR ORG
the United States GPE
37-40 CARDINAL
around 30 CARDINAL
CDC ORG
COVID-19 ORG
PCR ORG
Times ORG
CDC ORG
PCR ORG
33 CARDINAL
clinicians NORP
PCR ORG
CDC ORG
five CARDINAL
July 13, 2020 DATE
40 DATE
September 28, 2020 DATE
Clinical Infectious Diseases ORG
Jaafar GPE
al. GPE
nearly 4,000 CARDINAL

30 CARDINAL
CDC ORG
PCR ORG
December 1, 2020 DATE
40 CARDINAL
the New York Times ORG
CDC ORG
COVID-19 ORG
Fauci PERSON
the Mendacious Midget ORG
August DATE
one CARDINAL
Fauci PERSON
late January DATE
one CARDINAL
Fauci ORG
Today DATE
an hour TIME
Joe Biden PERSON
WHO ORG
PCR ORG
Diagnostic LOC
1 CARDINAL
NAT ORG
one hour TIME
Biden PERSON
WHO ORG
PCR ORG
Andy Swan PERSON
January 20, 2021 DATE
Fauci PERSON
July DATE
35 CARDINAL
35 DATE
Jaafar NORP
Florida GPE
December DATE
WHO ORG
PCR ORG
PCR ORG
higher than 30 CARDINAL
the New York Times ORG
Jaafar NORP
CNN ORG
Times ORG
August DATE
approximately 2.43 million CARDINAL
24.3 million CARDINAL
COVID-19 ORG
September DATE

Joe Biden PERSON
Anthony Fauci PERSON
a year DATE
Andy Swan PERSON
January 20, 2021 DATE
this season DATE
COVID ORG
Democrat NORP
Anti-Trump Postal Workers ORG
Garbage ORG
UK GPE
European Union ORG
British NORP
50/50 CARDINAL
Joe Biden PERSON
Gohmert Responds PERSON
May 2018 DATE
fourth ORDINAL
the month of May DATE
NOAA ORG
1880 DATE
March-May DATE
fourth ORDINAL
monthly DATE
NOAA ORG
National Centers for Environmental Information ORG
NOAA ORG
May 2018 DATE
May DATE
1.44 CARDINAL
F PRODUCT
the 20th century DATE
58.6 CARDINAL
F PRODUCT
fourth ORDINAL
May DATE
1880-2018 DATE
years 2014-2018 DATE
five CARDINAL
2016 CARDINAL
May DATE
1.58 CARDINAL
May 2018 DATE
the 42nd consecutive May DATE
the 401st consecutive month DATE
the 20th century DATE
North America LOC
Europe LOC
Asia LOC
Atlantic LOC

Pacific LOC
Indian Oceans NORP
May DATE
Canada GPE
Atlantic Ocean LOC
Greenland GPE
May DATE
2.05 CARDINAL
the 20th century DATE
52.0 CARDINAL
F PRODUCT
May DATE
2011 DATE
2013 DATE
seventh ORDINAL
May DATE
139-year DATE
May 2018 DATE
U.S. GPE
Europe LOC
5.4 CARDINAL
Canada GPE
Russia GPE
5.4 CARDINAL
Europe LOC
May DATE
continental ORG
1910 DATE
4.97 CARDINAL
2003 DATE
1.66 CARDINAL
F PRODUCT
May 2018 DATE
first ORDINAL
May DATE
3.6 CARDINAL
F PRODUCT
2.0 CARDINAL
European NORP
May DATE
Asia LOC
May DATE
2009 DATE
May DATE
1.19 CARDINAL
the 20th century DATE
monthly DATE
61.3 CARDINAL
fourth ORDINAL
May DATE

years 2014-2018 DATE
five CARDINAL
2016 CARDINAL
May DATE
1.39 CARDINAL
May 2018 DATE
May DATE
Arctic LOC
second ORDINAL
40-year DATE
420,000 square miles QUANTITY
8.1 percent PERCENT
1981-2010 DATE
the National Snow and Ice Data Center(link ORG
NOAA ORG
NASA ORG
May Arctic LOC
2016 DATE
Arctic LOC
Bering GPE
Chukchi Seas PERSON
May DATE
330,000 square miles QUANTITY
8.6 percent PERCENT
1981-2010 DATE
third ORDINAL
May DATE
Antarctic sea LOC
May DATE
the Weddell Sea LOC
Seasonal DATE
March-May 2018 DATE
March-May DATE
1.48 CARDINAL
the 20th century DATE
56.7 CARDINAL
F PRODUCT
fourth ORDINAL
March-May DATE
1880-2018 DATE
0.45 CARDINAL
F cooler PRODUCT
2016 DATE
March-May DATE
2014 DATE
three-month DATE
Europe LOC
the Middle East LOC
Africa LOC
China GPE

Mongolia GPE
South America LOC
Atlantic LOC
Indian NORP
Pacific Oceans NORP
March-May 2018 DATE
March-May DATE
2.34 CARDINAL
the 20th century DATE
46.4 CARDINAL
F PRODUCT
fifth ORDINAL
March-May DATE
139-year DATE
Five CARDINAL
March-May 2018 DATE
nine CARDINAL
continental ORG
1910 DATE
Europe LOC
South America LOC
third ORDINAL
March-May DATE
March-May DATE
1.15 CARDINAL
the 20th century DATE
61.0 CARDINAL
F PRODUCT
fourth ORDINAL
March-May DATE
Year-to-Date DATE
January-May 2018 DATE
1.39 CARDINAL
the 20th century DATE
55.5 CARDINAL
F PRODUCT
2010 DATE
fourth ORDINAL
January-May DATE
139-year DATE
2018 year DATE
0.59 CARDINAL
2016 DATE
2.18 CARDINAL
the 20th century DATE
42.8 CARDINAL
F PRODUCT
2002 DATE
fifth ORDINAL
January-May DATE

1.08 CARDINAL
the 20th century DATE
60.8 CARDINAL
fifth ORDINAL
January-May DATE
1880-2018 DATE
CDC Latest Update On Antibody Testing Results ORG
Rick Perry PERSON
U.S. GPE
Iran GPE
Americans NORP
Chelsea Clinton PERSON
18th ORDINAL
Veterans Mental Health And Suicide Prevention ORG
almost two thirds CARDINAL
Britons GPE
2,500 MONEY
Banned From Driving For Smoking ORG
Canada GPE
Ohio GPE
NHS ORG
Santiam Correctional Institution ORG
Santiam Correctional Institution ORG
The Washington Post ORG
Britain GPE
111 CARDINAL
NHS ORG
Yahaya Bello PERSON
Wingard PERSON
House ORG
Republicans NORP
Unveil Jobs Package ORG
Putin's End-of-Year Press Conference PERSON
Quotes GPE
Kremlin ORG
German Clinic's ORG
Diagnosis of Navalny's Condition ORG
NHS ORG
Global Warming PERSON
Jill Biden PERSON
10,000x CARDINAL
200 CARDINAL
NHS ORG
Two CARDINAL
Georgia GPE
Obama GPE
Trump ORG
COVID-19 ORG
10,000x CARDINAL
Hillary Clinton Boards PERSON

The Climate Crisis Train To Nowhere ORG
Milwaukee GPE
fourth ORDINAL
Kremlin ORG
German Clinic's ORG
Diagnosis of Navalny's Condition ORG
Pennsylvania GPE
Files Memorandum PERSON
2020 DATE
Marine Corps ORG
The Mexico Border Would Affect ORG
The Fight for Wisconsin's Soul - The New York Times ORG
Beef WORK_OF_ART
3 billion CARDINAL
up to half CARDINAL
50 years DATE
Neil Mackay's PERSON
Scotland GPE
2055 CARDINAL
2030s DATE
Caucasian Jack PERSON
Greenland GPE
annual DATE
Ministry of Ayurveda, Yoga & Naturopathy ORG
Unani PERSON
Siddha PERSON
Homoeopathy PERSON
Ayurveda ORG
Siddha PERSON
United Airlines ORG
Google ORG
almost two thirds CARDINAL
Britons GPE
Ballots PERSON
Michigan GPE
Dems NORP
B.C. GPE
ALS ORG
years DATE
24-hour TIME
N.Y. GPE
Competitive Elections Wisconsin ORG
State Assembly ORG
Islamic NORP
Exclusive: The TCF Center Election Fraud — Newly Discovered Video
Shows Late Night Deliveries WORK_OF_ART
Tens of Thousands CARDINAL
8 Hours TIME
Court ORG
Ameen Jogee ORG

Paul Stekler PERSON
at least 21 CARDINAL
Ted Cruz PERSON
Trump ORG
Paris GPE
Biden Quietly Pardons ORG
2,617 CARDINAL
Tory NORP
Dominic Raab PERSON
Iran GPE
Erbil ORG
Iraq GPE
Duncan Smith PERSON
Biden PERSON
COVID-19 Vaccination Efforts ORG
CDC ORG
the American Academy of Pediatrics ORG
Andrew Zimmerman PERSON
Kennedy Krieger Institute ORG
Johns Hopkins University ORG
third ORDINAL
years DATE
Zimmerman PERSON
Zimmerman PERSON
more than five thousand CARDINAL
2009 DATE
the late 1990s DATE
the Kennedy Krieger Institute ORG
Baltimore GPE
Zimmerman PERSON
Jon Poling PERSON
2000 DATE
nineteen-month-old DATE
Hannah PERSON
Hannah PERSON
Zimmerman PERSON
Richard Kelley PERSON
Kennedy Krieger's PERSON
Kennedy Krieger PERSON
Hannah PERSON
Georgetown GPE
MD GPE
Teri PERSON
Hannah PERSON
the next few years DATE
Zimmerman PERSON
2008 DATE
Hannah Poling PERSON
Zimmerman PERSON
Zimmerman PERSON

Zimmerman PERSON
Hannah PERSON
CDC ORG
the Department of Justice ORG
twenty-million-dollar MONEY
Polings FAC
2010 DATE
Hannah PERSON
Hannah Poling PERSON
Zimmerman PERSON
Kelley PERSON
the Department of Justice ORG
Drs PERSON
Zimmerman PERSON
Kelley PERSON
first ORDINAL
more than thirty years DATE
Tennessee GPE
Yates Hazlehurst PERSON
one CARDINAL
three CARDINAL
2009 DATE
only a year DATE
Poling ORG
Zimmerman.3 Drs PERSON
Zimmerman PERSON
Kelley PERSON
Hazlehurst NORP
Yates GPE
2009 DATE
the Omnibus Autism Proceeding ORG
Zimmerman PERSON
more than five thousand CARDINAL
2017 DATE
Drs PERSON
Kelley PERSON
Zimmerman PERSON
Yates GPE
The "Vaccine Court ORG
the National Childhood Vaccine Injury Act ORG
1986 DATE
the United States Court of Federal Claims ORG
American NORP
the Department of Health and Human Services ORG
HHS ORG
Department of Justice ORG
1989 DATE
more than $3.8 billion MONEY
Americans NORP
1998 DATE

twelve hundred CARDINAL
eighteen thousand CARDINAL
DTP ORG
MMR ORG
second ORDINAL
approximately 34 percent PERCENT
2017 DATE
just over $282 million MONEY
Rolf Hazlehurst PERSON
Tennessee GPE
Yates PERSON
the US Congress ORG
2013 DATE
Rolf Hazlehurst ORG
Vaccine PERSON
Discovery PRODUCT
American NORP
Vos PERSON
Fitzgerald PERSON
Evers PERSON
Wisconsin GPE
Honduran NORP
Britain GPE
Windrush PERSON
SNUBBED ORG
week DATE
Commonwealth ORG
Foreign Ministry ORG
Maria Zakharova PERSON
Moscow GPE
April 4, 2019 DATE
Biden Promises To Outlaw Gas-Powered Cars If Elected ORG
Michigan GPE
Kamala Harris PERSON
United States GPE
Democrats NORP
Social Security and Military ORG
Greenland GPE
annual DATE
Almost half CARDINAL
2001 DATE
AOC Proposes Ending Social Security Benefits After Age 75 ORG
British Medical Journal ORG
2007 DATE
67-year-old DATE
Dieneke Ferguson PERSON
three CARDINAL
15 months DATE
Three CARDINAL
four CARDINAL

the U.K.'s Daily Mail ORG
three days DATE
two CARDINAL
Four years DATE
daily DATE
eight grams QUANTITY
Myeloma PERSON
second ORDINAL
five years DATE
BMJ ORG
Ferguson ORG
Jamie Cavenagh PERSON
this day DATE
millions CARDINAL
Turmeric PRODUCT
2016 DATE
the Journal of Medicinal Food ORG
Ferguson ORG
Ferguson ORG
Fastest Rate FAC
Last 28 Centuries DATE
36% PERCENT
Megyn Kelly's ORG
Sin PERSON
Trump ORG
McCain ORG
Evans PERSON
Dueker PERSON
St. Louis GPE
Bill Gates PERSON
COVID-19 Vaccine PERSON
Nearly 1,000,000 CARDINAL
Kimchi May Have Helped Lower PERSON
COVID-19 Fatalities ORG
European NORP
Lisa Page PERSON
DNC Server Was Not Hacked ORG
Russia GPE
Swedish NORP
Photos Flood Social ORG
Zakharova ORG
New Salisbury Claims Aim to Distract Public From the Incident ORG
ENCYCLICAL LETTER LAUDATO SI' OF THE HOLY FATHER FRANCIS ON CARE FOR
OUR COMMON HOME ORG
Northern Ireland's GPE
Covid NORP
Rick Scott PERSON
Democrats NORP
Liberal States GPE
almost two thirds CARDINAL

```
Britons GPE
Masari NORP
USA GPE
World Cup EVENT
Netherlands GPE
2 CARDINAL
UK GPE
two-year DATE
Joe Biden PERSON
Rhode Island General Assembly ORG
Buhari Never PERSON
```

Exemple d'affichage graphique des entités nommées d'un texte

```python
nlp = spacy.load("en_core_web_sm")
text = dftrain['title'].iloc[3]  # Choisir le premier texte comme
exemple
doc = nlp(text)

displacy.render(doc, style="ent", jupyter=True)

<IPython.core.display.HTML object>
```

Cette fonction a pour but d'ajouter le type d'entité de chaque token détecté à coté de chaque entité détéctée dans le texte qu'elle prend en entrée, et retourne le texte modifié avec les informations ajoutées sur les entités nommées.

```python
def add_entity_name(text):
    if text is None or pd.isna(text):
        return ""

    doc = nlp(text)
    # Créer une liste pour stocker les nouveaux tokens
    nouveaux_tokens = []

    # Parcourir les tokens et ajouter des informations selon le type
d'entité
    for token in doc:
        if token.ent_type_:
            # Ajouter le nom de l'entité (type d'entité) à côté de
l'entité
            nouveaux_tokens.append(f"{token.text}
({token.ent_type_})")
        else:
            nouveaux_tokens.append(token.text)

    # Reconstruire le texte avec les informations ajoutées
    nouveau_texte = " ".join(nouveaux_tokens)
    return nouveau_texte
```

Application de la fonction précédente à la colonne "title" avec la méthode apply

```python
dftrain['title'] = dftrain['title'].astype(str)

# application de la fonction à la colonne "title"
dftrain['title'] = dftrain['title'].apply(add_entity_name)

print(dftrain['title'])
```

```
847      Principal banned candy canes because ' J ' sha...
447      On the Capitol (FAC) : Part - time pay for par...
2243     One (CARDINAL) of the strengths — and weakness...
1202     Australia (GPE) 's Great (WORK_OF_ART) Barrier...
1059     Hospitals deluged by 5,000 (CARDINAL) diabetic...
                              ...
637      UK (GPE) inflation soars to two (DATE) - (DATE...
1022                                                   nan
1515     BREAKING EXCLUSIVE : System ' Glitch ' Also Un...
1432     State of Rhode (ORG) Island (ORG) General (ORG...
628      President Buhari (PERSON) Never (PERSON) Said ...
Name: title, Length: 1900, dtype: object
```

On scinde les données de la colonne "title" en jeu d'apprentissage et jeu de test

```python
X=dftrain["title"]
print(X)

X_train,X_test,y_train,y_test=train_test_split(X,y,test_size =
0.2,random_state=8)
print("X_train is",X_train)
print("y_train is",y_train)
print("X_test is",X_test)
print("y_test is",y_test)
```

```
847      Principal banned candy canes because ' J ' sha...
447      On the Capitol (FAC) : Part - time pay for par...
2243     One (CARDINAL) of the strengths — and weakness...
1202     Australia (GPE) 's Great (WORK_OF_ART) Barrier...
1059     Hospitals deluged by 5,000 (CARDINAL) diabetic...
                              ...
637      UK (GPE) inflation soars to two (DATE) - (DATE...
1022                                                   nan
1515     BREAKING EXCLUSIVE : System ' Glitch ' Also Un...
1432     State of Rhode (ORG) Island (ORG) General (ORG...
628      President Buhari (PERSON) Never (PERSON) Said ...
Name: title, Length: 1900, dtype: object
X_train is 1625    You Can be Fined $ 2,500 (MONEY) And Banned (O...
2164              - The (ORG) Washington (ORG) Post (ORG)
221      Why Does Fauci (PERSON) Hold (PERSON) Patents ...
721      A discussion of ' smokers ' black lungs ' star...
2413     We wo n't let women be second (ORDINAL) class ...
                              ...
```

```
840      USC (ORG) Professor Joel (PERSON) Hay (PERSON)...
1155     When Will The (WORK_OF_ART) Planet (WORK_OF_AR...
453      United (ORG) Airlines (ORG) Flight Attendant S...
2146     Union blasts Gov. Scott (PERSON) Walker (PERSO...
948      A Report on Updated Overtime and Food Expenses...
Name: title, Length: 1520, dtype: object
y_train is 1625      TRUE/FALSE
2164    MIXTURE/OTHER
221         TRUE/FALSE
721         TRUE/FALSE
2413    MIXTURE/OTHER
            ...
840     MIXTURE/OTHER
1155    MIXTURE/OTHER
453         TRUE/FALSE
2146        TRUE/FALSE
948     MIXTURE/OTHER
Name: regrouped, Length: 1520, dtype: object
X_test is 1847      One (CARDINAL) of the strengths – and weakness...
103      Alabama (ORG) Supreme (ORG) Court (ORG) Recogn...
818      Central (LOC) Europe (LOC) would veto any Brex...
68       Banned By Facebook – DO NOT READ ! Bad (PERSON...
2039     Early help for children cut by half (CARDINAL)...
                          ...
1461     Ten (TIME) minutes (TIME) ' exercise each (DAT...
879                                                    nan
606      Abe (PERSON) Lincoln (PERSON) Statue Vandalize...
859      Australian (NORP) Authorities : Arsonists to B...
2268     Florida (GPE) GOP (ORG) candidates for governo...
Name: title, Length: 380, dtype: object
y_test is 1847      TRUE/FALSE
103         TRUE/FALSE
818         TRUE/FALSE
68          TRUE/FALSE
2039    MIXTURE/OTHER
            ...
1461        TRUE/FALSE
879     MIXTURE/OTHER
606     MIXTURE/OTHER
859         TRUE/FALSE
2268    MIXTURE/OTHER
Name: regrouped, Length: 380, dtype: object
```

Ici, c'est une étape importante, on va tester différents classifieurs, pour chacun des classifieurs, on va appliquer le prétraitement + Vectorisation TfIdf, et on applique une cross_val_score avec un Kfold de 10 fois, par la suite on stocke dans une liste all_results la moyenne des accuracy + l'écart type et on la trie par ordre décroissant de moyenne d'accuracy et d'écart type. on remarque que les 2 meilleurs sont SVM et RF qu'on va séléctionner pour leur appliquer le GridSearch sur les paramètres des prétraitements + leurs hyperparamètres pour pouvoir choisir le meilleur.

```python
# Utilisez la méthode ravel() pour transformer y_train en un tableau
unidimensionnel
X_train = np.ravel(X_train)

np.random.seed(42)  # Set the random seed for NumPy

score = 'accuracy'
seed = 7
allresults = []
results = []
names = []


# Liste des modèles à tester
models = [
    ('MultinomialNB', MultinomialNB()),
    ('LogisticRegression', LogisticRegression(random_state=42))
]


models.append(('KNN', KNeighborsClassifier()))
models.append(('CART', DecisionTreeClassifier(random_state=42)))
models.append(('RF', RandomForestClassifier(random_state=42)))
models.append(('SVM', SVC(random_state=42)))

# Création d'un pipeline pour chaque modèle
pipelines = []
for name,model in models:
    pipeline = Pipeline([
        ('normalize', TextNormalizer()),
        ('tfidf', TfidfVectorizer()),
        (name,model)
    ])
    pipelines.append((name,pipeline))
    #pipeline.fit(X_train,y_train)
all_results=[]
scores=[]
for p in pipelines:
    print(p[1])
    # cross validation en 10 fois
    kfold = KFold(n_splits=10,random_state=seed,shuffle=True)

#    print ("Evaluation de ",p)
    start_time = time.time()
    # application de la classification
    cv_results = cross_val_score(p[1],X_train,y_train, cv=kfold,
scoring=score)
    #print("Pour le classifieur",p[0],"on a un score
```

```python
de",cv_results.mean(),"et un écart type de",cv_results.std())
    scores.append(cv_results)
    names.append(p[0])

    all_results.append((p[0],cv_results.mean(),cv_results.std()))
    end_time = time.time()



print("all resultats", all_results)

all_results = sorted(all_results, key=lambda x: (-x[1], -x[2]))
print("all resultats", all_results)
    # affichage des résultats
#print ('\nLe meilleur resultat : ',max(results))

Pipeline(steps=[('normalize', TextNormalizer()), ('tfidf',
TfidfVectorizer()),
                ('MultinomialNB', MultinomialNB())])
Pipeline(steps=[('normalize', TextNormalizer()), ('tfidf',
TfidfVectorizer()),
                ('LogisticRegression',
LogisticRegression(random_state=42))])
Pipeline(steps=[('normalize', TextNormalizer()), ('tfidf',
TfidfVectorizer()),
                ('KNN', KNeighborsClassifier())])
Pipeline(steps=[('normalize', TextNormalizer()), ('tfidf',
TfidfVectorizer()),
                ('CART', DecisionTreeClassifier(random_state=42))])
Pipeline(steps=[('normalize', TextNormalizer()), ('tfidf',
TfidfVectorizer()),
                ('RF', RandomForestClassifier(random_state=42))])
Pipeline(steps=[('normalize', TextNormalizer()), ('tfidf',
TfidfVectorizer()),
                ('SVM', SVC(random_state=42))])
all resultats [('MultinomialNB', 0.8, 0.023938691314026636),
('LogisticRegression', 0.7881578947368422, 0.025446157376070665),
('KNN', 0.6407894736842106, 0.049600202170975695), ('CART',
0.832236842105263, 0.03509114566603837), ('RF', 0.8342105263157895,
0.023129468199009143), ('SVM', 0.8605263157894736,
0.025784102555612393)]
all resultats [('SVM', 0.8605263157894736, 0.025784102555612393),
('RF', 0.8342105263157895, 0.023129468199009143), ('CART',
0.832236842105263, 0.03509114566603837), ('MultinomialNB', 0.8,
0.023938691314026636), ('LogisticRegression', 0.7881578947368422,
0.025446157376070665), ('KNN', 0.6407894736842106,
0.049600202170975695)]
```
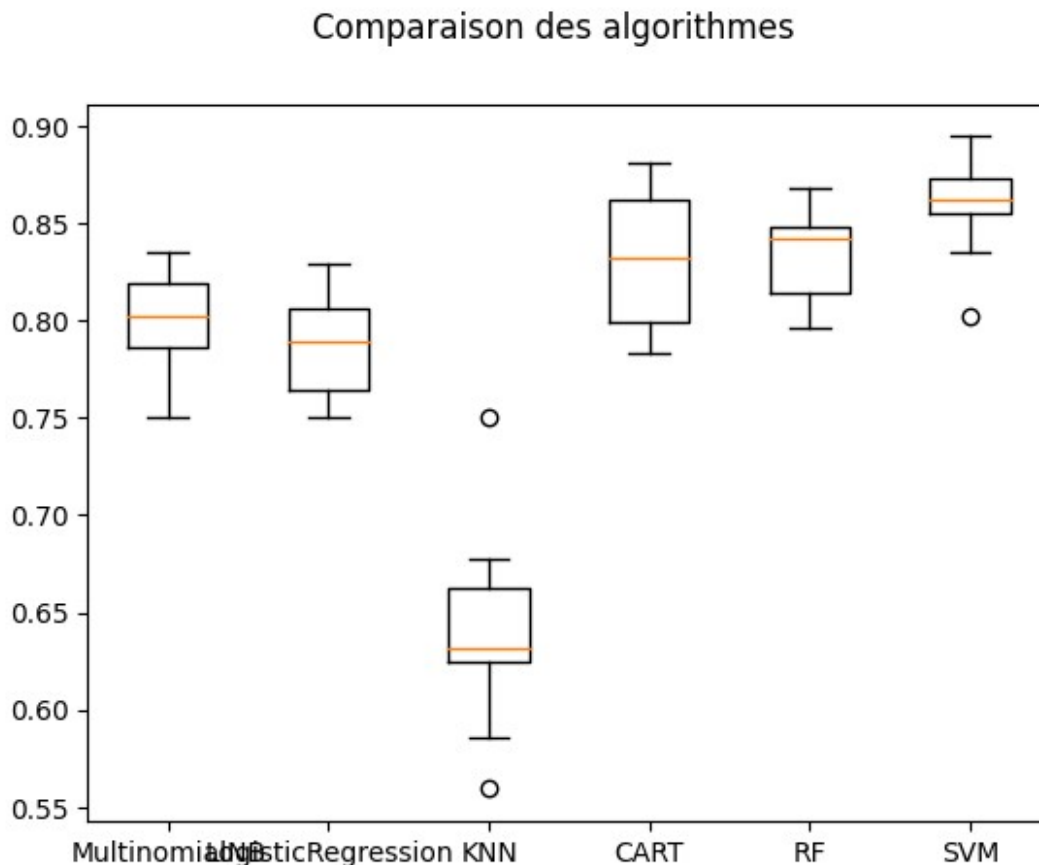
On affiche les boites à moustache pour mieux visualiser les résultats

```
fig = plt.figure()
fig.suptitle('Comparaison des algorithmes')
ax = fig.add_subplot(111)
plt.boxplot(scores)
ax.set_xticklabels(names)

[Text(1, 0, 'MultinomialNB'),
 Text(2, 0, 'LogisticRegression'),
 Text(3, 0, 'KNN'),
 Text(4, 0, 'CART'),
 Text(5, 0, 'RF'),
 Text(6, 0, 'SVM')]
```


Comparaison des algorithmes

On a un pipeline pour chaque prétraitement différent, on essaye pas mal (miniscule, lemmatisation, miniscule + lemmatisation..) et on stocke le fit_transorm de nos X_train, X_test sur les pipelines dans des listes qui vont contenir tous les fit_transform des pipelines pour chaque classifieur, par la suite on parcourt ces listes là, on itère dessus, et chaque élement de la liste (train) va passer par le GridSearch et puis on predict sur son corresapondant dans liste (test).

```
np.random.seed(42)  # Set the random seed for NumPy
print("y_train", y_train.shape)
```

```python
print("y_test", y_test.shape)
#X_test = np.ravel(X_test)
print("X_test", X_test.shape)
# le plus simple est de faire un test sur differents pipelines.
# pipeline de l'utilisation de CountVectorizer sur le texte avec
differents pre-traitements
CV_brut = Pipeline([('cleaner', TextNormalizer()),
                    ('count_vectorizer',
CountVectorizer(lowercase=False))])
CV_lowcase = Pipeline([('cleaner',
TextNormalizer(removestopwords=False,lowercase=True,

getstemmer=False,removedigit=False)),
                    ('count_vectorizer',
CountVectorizer(lowercase=False))])
CV_lowStop = Pipeline([('cleaner',
TextNormalizer(removestopwords=True,lowercase=True,

getstemmer=False,removedigit=False)),
                    ('count_vectorizer',
CountVectorizer(lowercase=False))])

CV_lowStopstem = Pipeline([('cleaner',
TextNormalizer(removestopwords=True,lowercase=True,

getstemmer=True,removedigit=False)),
                    ('count_vectorizer',
CountVectorizer(lowercase=False))])

# pipeline de l'utilisation de TfidfVectorizer avec differents pre-
traitements
TFIDF_brut = Pipeline ([('cleaner', TextNormalizer()),
                    ('tfidf_vectorizer',
TfidfVectorizer(lowercase=False))])

TFIDF_lowcase = Pipeline([('cleaner',
TextNormalizer(removestopwords=False,lowercase=True,

getstemmer=False,removedigit=False)),
                    ('tfidf_vectorizer',
TfidfVectorizer(lowercase=False))])
TFIDF_lowStop = Pipeline([('cleaner',
TextNormalizer(removestopwords=True,lowercase=True,

getstemmer=False,removedigit=False)),
                    ('tfidf_vectorizer',
TfidfVectorizer(lowercase=False))])

TFIDF_lowStopstem = Pipeline([('cleaner',
```

```python
TextNormalizer(removestopwords=True,lowercase=True,

getstemmer=True,removedigit=False)),
                    ('tfidf_vectorizer',
TfidfVectorizer(lowercase=False))])


# Liste de tous les modeles à tester
all_models = [
    ("CV_brut", CV_brut),
    ("CV_lowcase", CV_lowcase),
    ("CV_lowStop", CV_lowStop),
    ("CV_lowStopstem",CV_lowStopstem),
    ("TFIDF_lowcase", TFIDF_lowcase),
    ("TFIDF_lowStop", TFIDF_lowStop),
    ("TFIDF_lowStopstem",TFIDF_lowStopstem),
    ("TFIDF_brut", TFIDF_brut)
]

X_train_title_SVC = []
X_test_title_SVC = []

X_train_title_RandomForestClassifier = []
X_test_title_RandomForestClassifier = []


for name, pipeline in all_models :

X_train_title_SVC.append(pipeline.fit_transform(X_train).toarray())
    X_test_title_SVC.append(pipeline.transform(X_test).toarray())

X_train_title_RandomForestClassifier.append(pipeline.fit_transform(X_t
rain).toarray())

X_test_title_RandomForestClassifier.append(pipeline.transform(X_test).
toarray())




models = {
    'SVC': SVC(random_state=42),
    'RandomForestClassifier': RandomForestClassifier(random_state=42)
}


params = {'SVC': [{'C': [0.001, 0.01, 0.1, 1,2,5,7,10]},
            {'gamma': [0.001, 0.01, 0.1,0.2,0.3,0.5,0.7,1]},
            {'kernel': ['linear', 'rbf']}],
```

```python
    'RandomForestClassifier': [{'n_estimators': [10, 50, 100, 200,
300]},
                               {'max_features': ['auto', 'sqrt',
'log2']}]
}

for model_name, model in models.items():
    score='accuracy'
    X_train_title = eval('X_train_title_' + model_name)
    X_test_title = eval('X_test_title_' + model_name)
    for i in range (len(X_train_title)):
      grid_search = GridSearchCV(model, params[model_name], n_jobs=-1,
verbose=1,scoring=score)
      print("grid search fait")
      print("X_train",X_train_title[i].shape)
      print("y_train",y_train.shape)
      grid_search.fit(X_train_title[i],y_train)
      print ('meilleur score %0.3f'%(grid_search.best_score_),'\n')
      print ('meilleur estimateur',grid_search.best_estimator_,'\n')
      y_pred = grid_search.predict(X_test_title[i])
      MyshowAllScores(y_test,y_pred)

      print("Ensemble des meilleurs paramètres :")
      best_parameters = grid_search.best_estimator_.get_params()
      for param_dict in params[model_name]:
        for param_name, param_value in param_dict.items():
            print("\t%s: %r" % (param_name,
best_parameters[param_name]))
```

```
y_train (1520,)
y_test (380,)
X_test (380,)
grid search fait
X_train (1520, 8440)
y_train (1520,)
Fitting 5 folds for each of 18 candidates, totalling 90 fits
meilleur score 0.873

meilleur estimateur SVC(gamma=0.5, random_state=42)

Accuracy : 0.882
Classification Report
               precision    recall  f1-score   support

MIXTURE/OTHER    0.91824   0.82022   0.86647       178
   TRUE/FALSE    0.85520   0.93564   0.89362       202

     accuracy                        0.88158       380
    macro avg    0.88672   0.87793   0.88004       380
 weighted avg    0.88473   0.88158   0.88090       380
```

```
Ensemble des meilleurs paramètres :
      C: 1.0
      gamma: 0.5
      kernel: 'rbf'
grid search fait
X_train (1520, 6959)
y_train (1520,)
Fitting 5 folds for each of 18 candidates, totalling 90 fits
meilleur score 0.873

meilleur estimateur SVC(gamma=0.5, random_state=42)

Accuracy : 0.887
Classification Report
               precision     recall   f1-score    support

MIXTURE/OTHER     0.92994    0.82022    0.87164        178
   TRUE/FALSE     0.85650    0.94554    0.89882        202

     accuracy                           0.88684        380
    macro avg     0.89322    0.88288    0.88523        380
 weighted avg     0.89090    0.88684    0.88609        380

Ensemble des meilleurs paramètres :
      C: 1.0
      gamma: 0.5
      kernel: 'rbf'
grid search fait
X_train (1520, 6839)
y_train (1520,)
Fitting 5 folds for each of 18 candidates, totalling 90 fits
meilleur score 0.873

meilleur estimateur SVC(gamma=0.7, random_state=42)

Accuracy : 0.889
Classification Report
               precision     recall   f1-score    support

MIXTURE/OTHER     0.93590    0.82022    0.87425        178
   TRUE/FALSE     0.85714    0.95050    0.90141        202

     accuracy                           0.88947        380
    macro avg     0.89652    0.88536    0.88783        380
 weighted avg     0.89403    0.88947    0.88869        380

Ensemble des meilleurs paramètres :
      C: 1.0
```

```
      gamma: 0.7
      kernel: 'rbf'
grid search fait
X_train (1520, 5026)
y_train (1520,)
Fitting 5 folds for each of 18 candidates, totalling 90 fits
meilleur score 0.873

meilleur estimateur SVC(gamma=0.7, random_state=42)

Accuracy : 0.889
Classification Report
              precision    recall  f1-score   support

MIXTURE/OTHER    0.93590   0.82022   0.87425       178
   TRUE/FALSE    0.85714   0.95050   0.90141       202

     accuracy                        0.88947       380
    macro avg    0.89652   0.88536   0.88783       380
 weighted avg    0.89403   0.88947   0.88869       380

Ensemble des meilleurs paramètres :
      C: 1.0
      gamma: 0.7
      kernel: 'rbf'
grid search fait
X_train (1520, 6959)
y_train (1520,)
Fitting 5 folds for each of 18 candidates, totalling 90 fits
meilleur score 0.844

meilleur estimateur SVC(C=5, random_state=42)

Accuracy : 0.863
Classification Report
              precision    recall  f1-score   support

MIXTURE/OTHER    0.85795   0.84831   0.85311       178
   TRUE/FALSE    0.86765   0.87624   0.87192       202

     accuracy                        0.86316       380
    macro avg    0.86280   0.86228   0.86251       380
 weighted avg    0.86311   0.86316   0.86311       380

Ensemble des meilleurs paramètres :
      C: 5
      gamma: 'scale'
      kernel: 'rbf'
grid search fait
```

```
X_train (1520, 6839)
y_train (1520,)
Fitting 5 folds for each of 18 candidates, totalling 90 fits
meilleur score 0.848

meilleur estimateur SVC(C=2, random_state=42)

Accuracy : 0.876
Classification Report
               precision    recall  f1-score   support

MIXTURE/OTHER     0.86592   0.87079   0.86835       178
   TRUE/FALSE     0.88557   0.88119   0.88337       202

     accuracy                         0.87632       380
    macro avg     0.87575   0.87599   0.87586       380
 weighted avg     0.87637   0.87632   0.87634       380

Ensemble des meilleurs paramètres :
     C: 2
     gamma: 'scale'
     kernel: 'rbf'
grid search fait
X_train (1520, 5026)
y_train (1520,)
Fitting 5 folds for each of 18 candidates, totalling 90 fits
meilleur score 0.837

meilleur estimateur SVC(C=5, random_state=42)

Accuracy : 0.879
Classification Report
               precision    recall  f1-score   support

MIXTURE/OTHER     0.85106   0.89888   0.87432       178
   TRUE/FALSE     0.90625   0.86139   0.88325       202

     accuracy                         0.87895       380
    macro avg     0.87866   0.88013   0.87878       380
 weighted avg     0.88040   0.87895   0.87906       380

Ensemble des meilleurs paramètres :
     C: 5
     gamma: 'scale'
     kernel: 'rbf'
grid search fait
X_train (1520, 8440)
y_train (1520,)
Fitting 5 folds for each of 18 candidates, totalling 90 fits
```

```
meilleur score 0.851

meilleur estimateur SVC(C=1, random_state=42)

Accuracy : 0.861
Classification Report
               precision    recall  f1-score   support

MIXTURE/OTHER    0.84916   0.85393   0.85154       178
   TRUE/FALSE    0.87065   0.86634   0.86849       202

     accuracy                        0.86053       380
    macro avg    0.85990   0.86013   0.86001       380
 weighted avg    0.86058   0.86053   0.86055       380


Ensemble des meilleurs paramètres :
     C: 1
     gamma: 'scale'
     kernel: 'rbf'
grid search fait
X_train (1520, 8440)
y_train (1520,)
Fitting 5 folds for each of 8 candidates, totalling 40 fits
meilleur score 0.798

meilleur estimateur RandomForestClassifier(max_features='log2',
random_state=42)

Accuracy : 0.805
Classification Report
               precision    recall  f1-score   support

MIXTURE/OTHER    0.73853   0.90449   0.81313       178
   TRUE/FALSE    0.89506   0.71782   0.79670       202

     accuracy                        0.80526       380
    macro avg    0.81680   0.81116   0.80492       380
 weighted avg    0.82174   0.80526   0.80440       380


Ensemble des meilleurs paramètres :
     n_estimators: 100
     max_features: 'log2'
grid search fait
X_train (1520, 6959)
y_train (1520,)
Fitting 5 folds for each of 8 candidates, totalling 40 fits
meilleur score 0.811

meilleur estimateur RandomForestClassifier(max_features='log2',
```

```
random_state=42)

Accuracy : 0.842
Classification Report
               precision    recall  f1-score   support

MIXTURE/OTHER    0.77315   0.93820   0.84772       178
   TRUE/FALSE    0.93293   0.75743   0.83607       202

     accuracy                        0.84211       380
    macro avg    0.85304   0.84781   0.84189       380
 weighted avg    0.85808   0.84211   0.84152       380

Ensemble des meilleurs paramètres :
     n_estimators: 100
     max_features: 'log2'
grid search fait
X_train (1520, 6839)
y_train (1520,)
Fitting 5 folds for each of 8 candidates, totalling 40 fits
meilleur score 0.801

meilleur estimateur RandomForestClassifier(max_features='log2',
random_state=42)

Accuracy : 0.834
Classification Report
               precision    recall  f1-score   support

MIXTURE/OTHER    0.75785   0.94944   0.84289       178
   TRUE/FALSE    0.94268   0.73267   0.82451       202

     accuracy                        0.83421       380
    macro avg    0.85026   0.84106   0.83370       380
 weighted avg    0.85610   0.83421   0.83312       380

Ensemble des meilleurs paramètres :
     n_estimators: 100
     max_features: 'log2'
grid search fait
X_train (1520, 5026)
y_train (1520,)
Fitting 5 folds for each of 8 candidates, totalling 40 fits
meilleur score 0.809

meilleur estimateur RandomForestClassifier(max_features='log2',
random_state=42)

Accuracy : 0.855
```

```
Classification Report
              precision    recall  f1-score   support

MIXTURE/OTHER     0.79147   0.93820   0.85861       178
   TRUE/FALSE     0.93491   0.78218   0.85175       202

     accuracy                         0.85526       380
    macro avg     0.86319   0.86019   0.85518       380
 weighted avg     0.86772   0.85526   0.85497       380

Ensemble des meilleurs paramètres :
     n_estimators: 100
     max_features: 'log2'
grid search fait
X_train (1520, 6959)
y_train (1520,)
Fitting 5 folds for each of 8 candidates, totalling 40 fits
meilleur score 0.810

meilleur estimateur RandomForestClassifier(n_estimators=50,
random_state=42)

Accuracy : 0.821
Classification Report
              precision    recall  f1-score   support

MIXTURE/OTHER     0.75701   0.91011   0.82653       178
   TRUE/FALSE     0.90361   0.74257   0.81522       202

     accuracy                         0.82105       380
    macro avg     0.83031   0.82634   0.82087       380
 weighted avg     0.83494   0.82105   0.82052       380

Ensemble des meilleurs paramètres :
     n_estimators: 50
     max_features: 'sqrt'
grid search fait
X_train (1520, 6839)
y_train (1520,)
Fitting 5 folds for each of 8 candidates, totalling 40 fits
meilleur score 0.806

meilleur estimateur RandomForestClassifier(n_estimators=300,
random_state=42)

Accuracy : 0.811
Classification Report
              precision    recall  f1-score   support
```

```
MIXTURE/OTHER      0.74091    0.91573    0.81910         178
   TRUE/FALSE      0.90625    0.71782    0.80110         202

     accuracy                           0.81053         380
    macro avg      0.82358    0.81678    0.81010         380
 weighted avg      0.82880    0.81053    0.80953         380
```

Ensemble des meilleurs paramètres :
     n_estimators: 300
     max_features: 'sqrt'
grid search fait
X_train (1520, 5026)
y_train (1520,)
Fitting 5 folds for each of 8 candidates, totalling 40 fits
meilleur score 0.806

meilleur estimateur RandomForestClassifier(max_features='log2',
random_state=42)

Accuracy : 0.866
Classification Report

```
                 precision    recall   f1-score    support

MIXTURE/OTHER      0.81281    0.92697    0.86614         178
   TRUE/FALSE      0.92655    0.81188    0.86544         202

     accuracy                           0.86579         380
    macro avg      0.86968    0.86942    0.86579         380
 weighted avg      0.87327    0.86579    0.86577         380
```

Ensemble des meilleurs paramètres :
     n_estimators: 100
     max_features: 'log2'
grid search fait
X_train (1520, 8440)
y_train (1520,)
Fitting 5 folds for each of 8 candidates, totalling 40 fits
meilleur score 0.799

meilleur estimateur RandomForestClassifier(n_estimators=300,
random_state=42)

Accuracy : 0.837
Classification Report

```
                 precision    recall   f1-score    support

MIXTURE/OTHER      0.77619    0.91573    0.84021         178
   TRUE/FALSE      0.91176    0.76733    0.83333         202
```

```
    accuracy                              0.83684      380
   macro avg      0.84398   0.84153     0.83677      380
weighted avg      0.84826   0.83684     0.83655      380
```

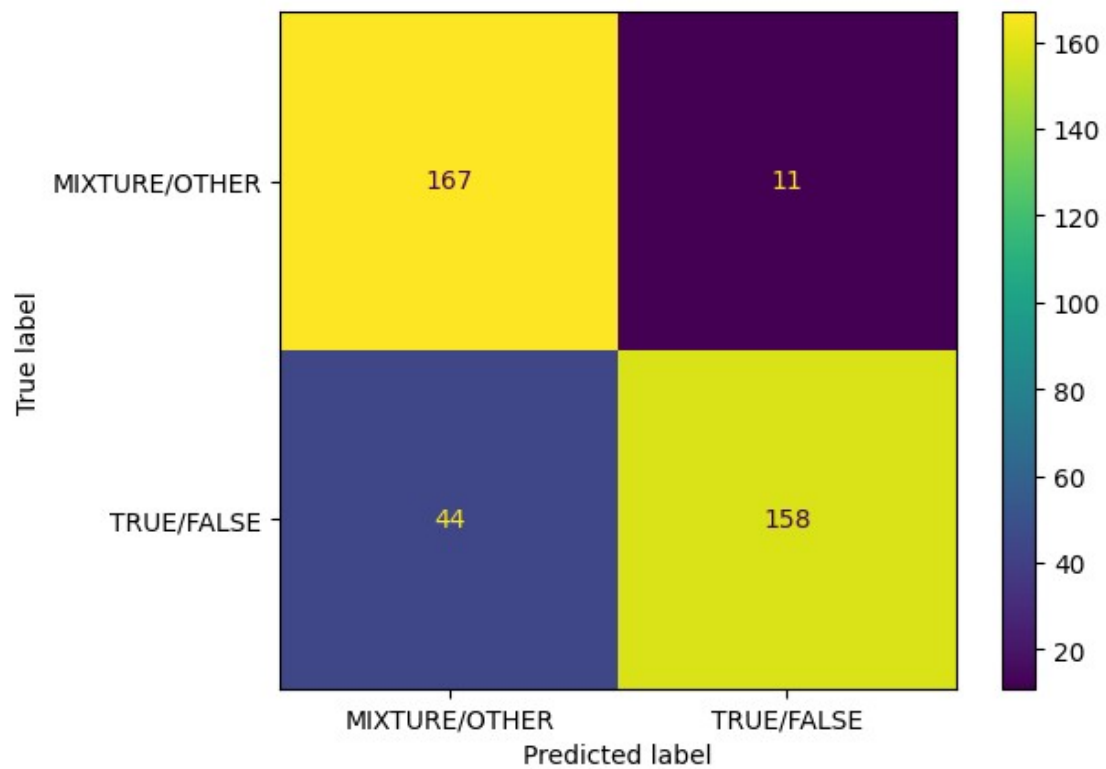Ensemble des meilleurs paramètres :
    n_estimators: 300
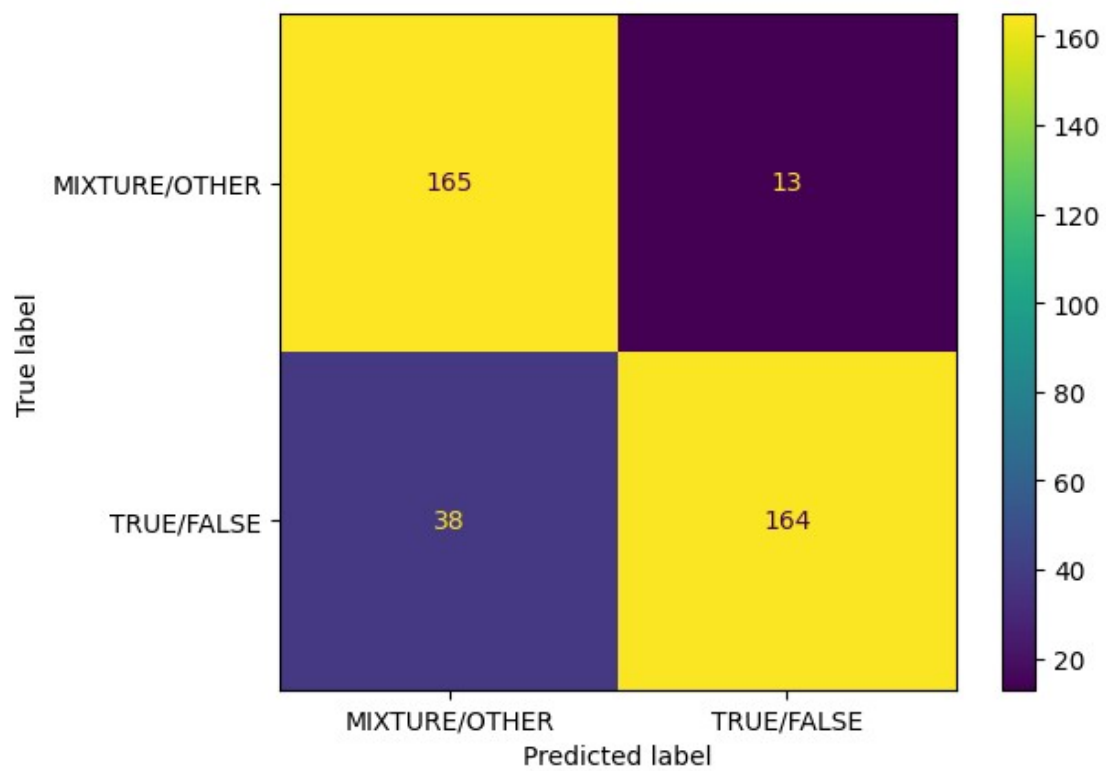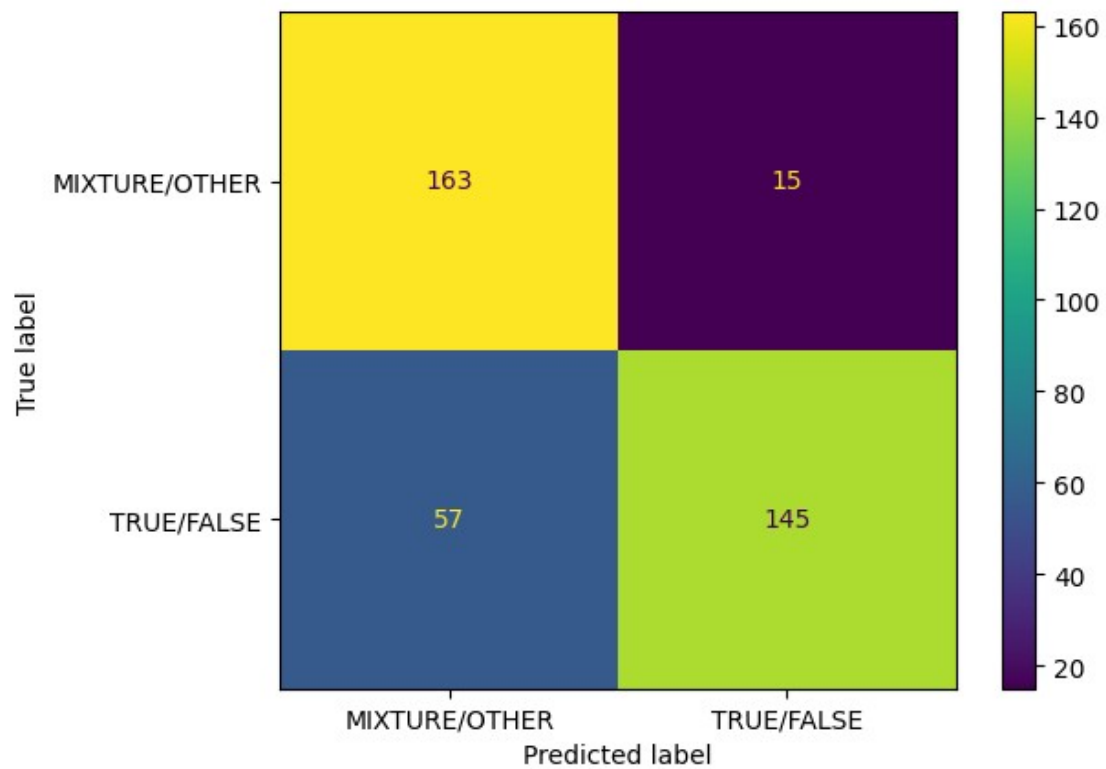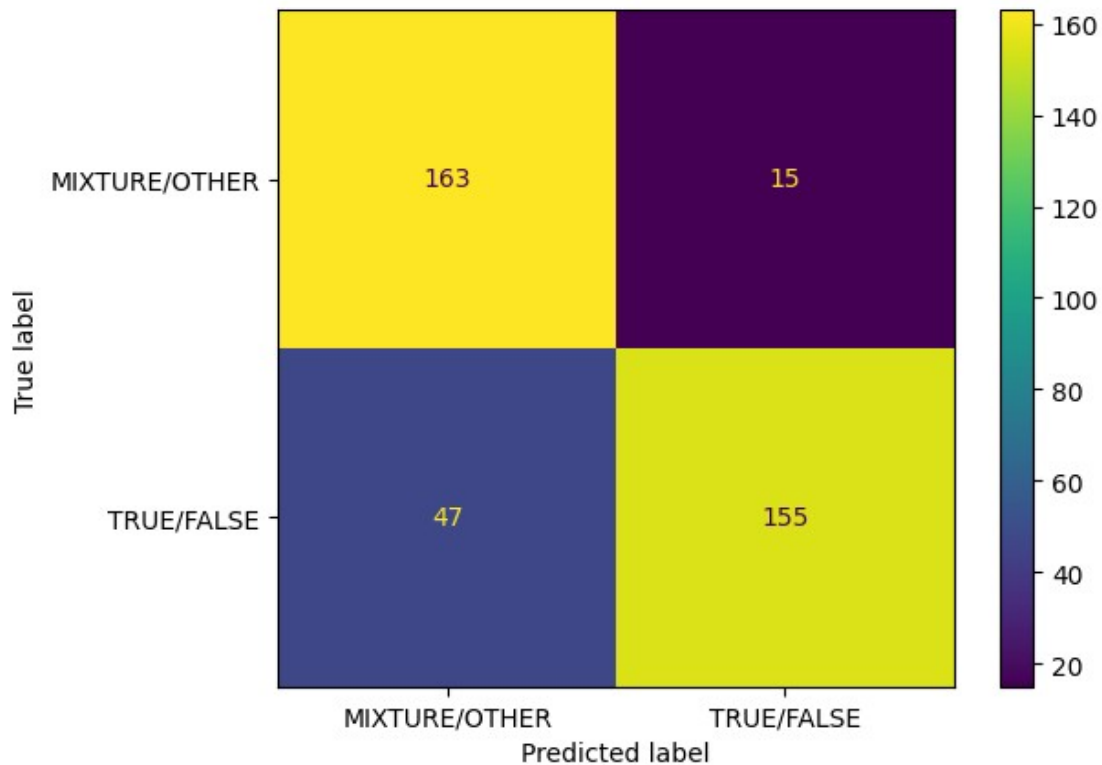   max_features: 'sqrt'
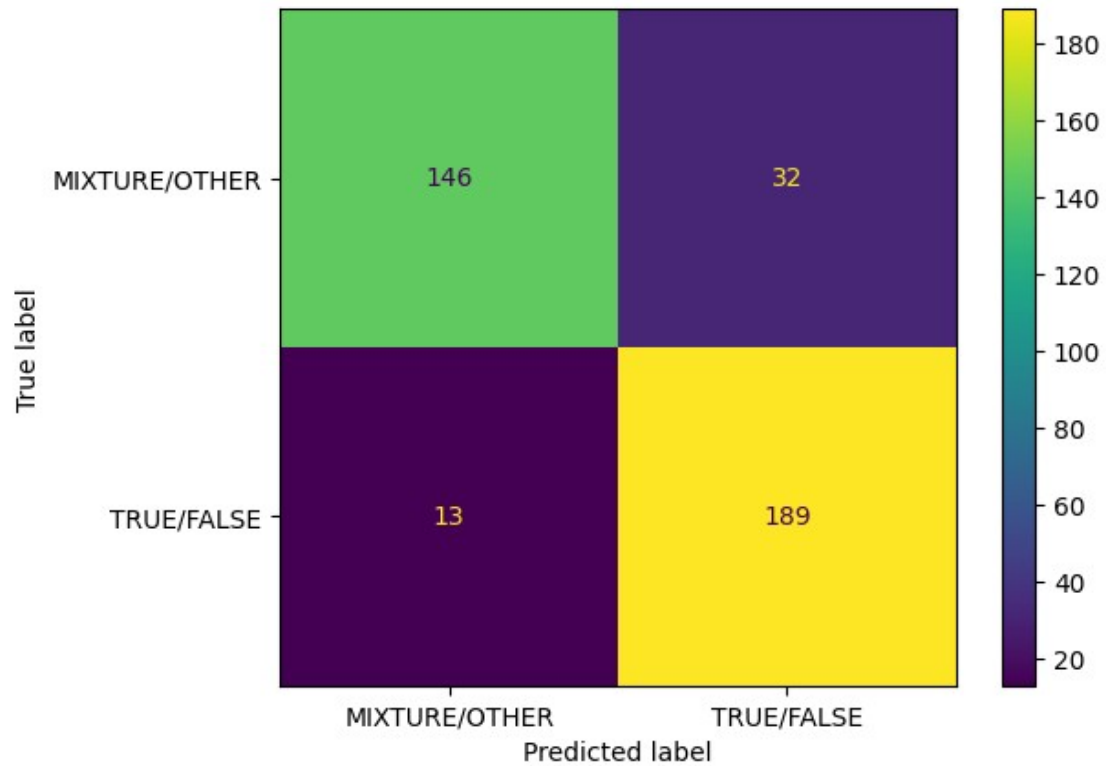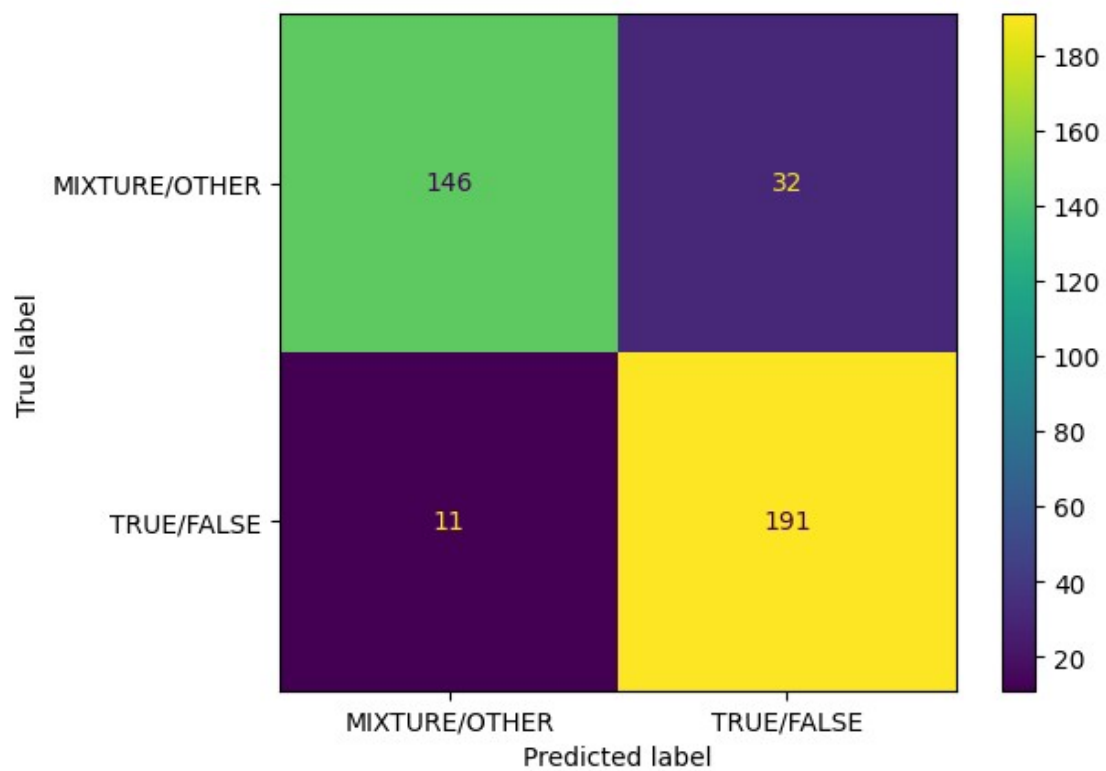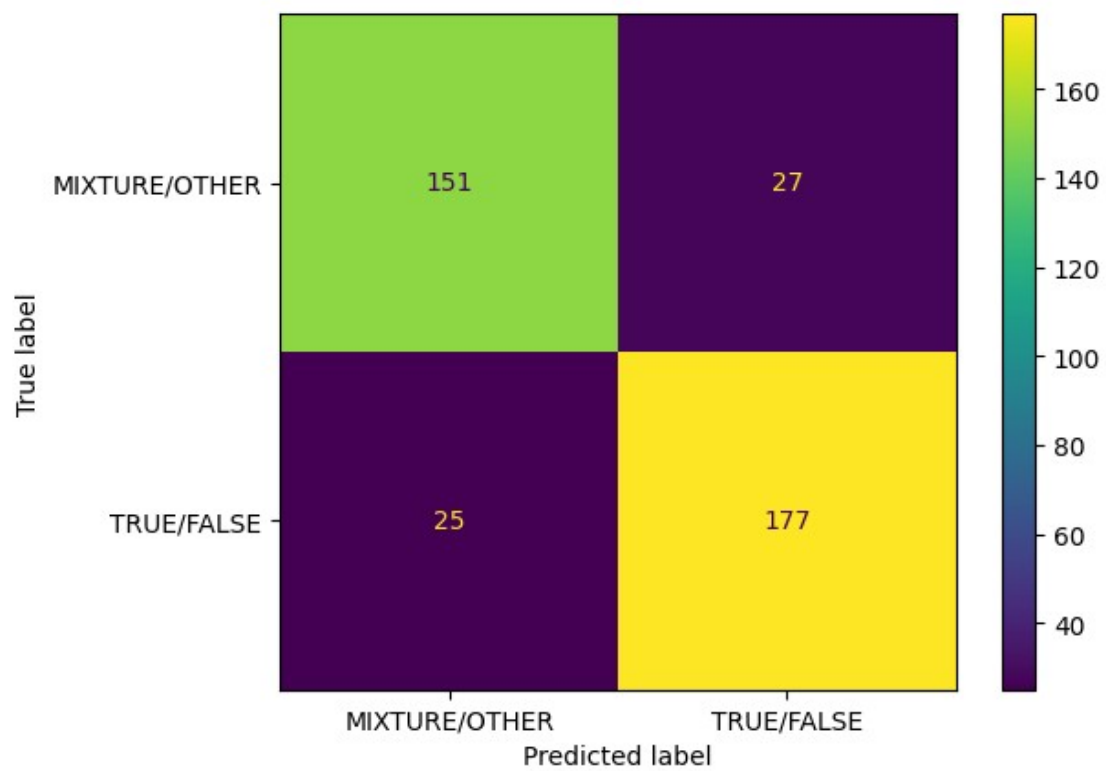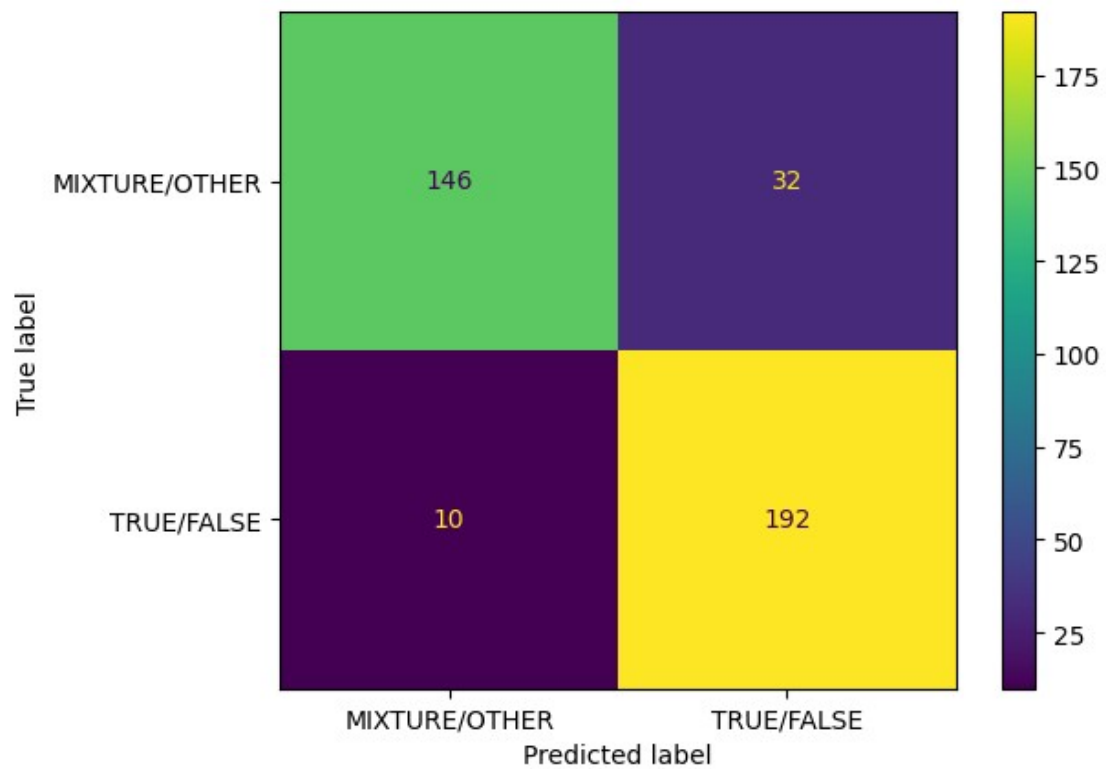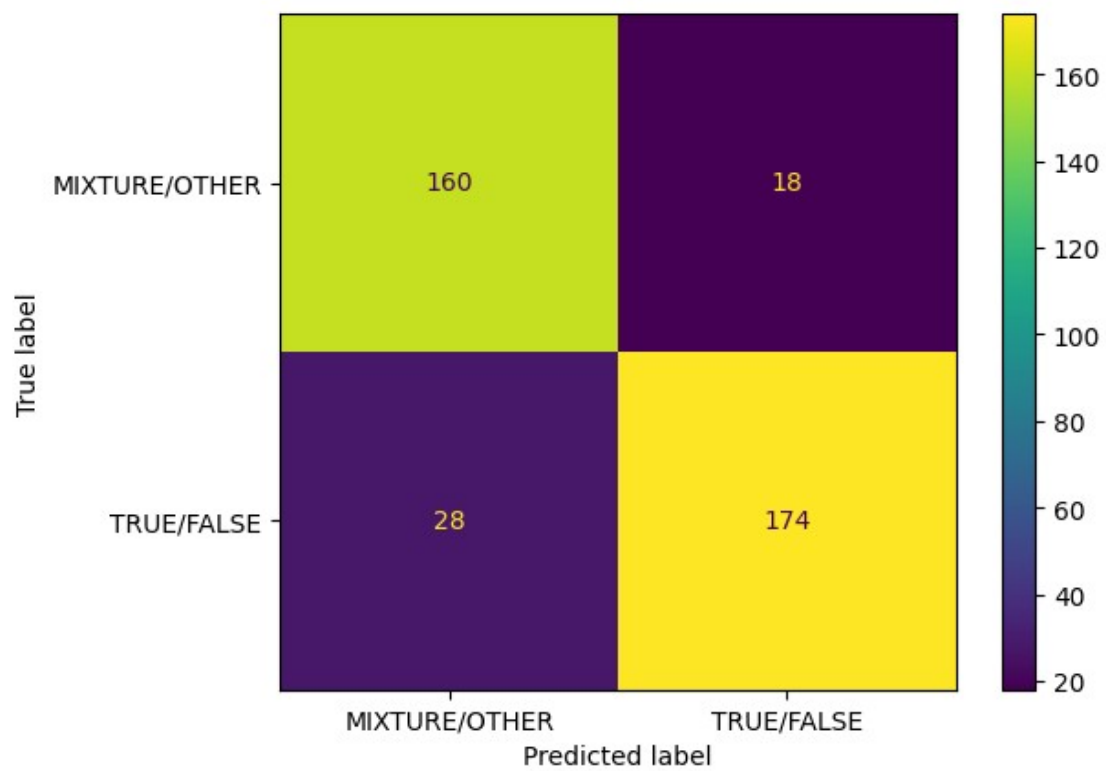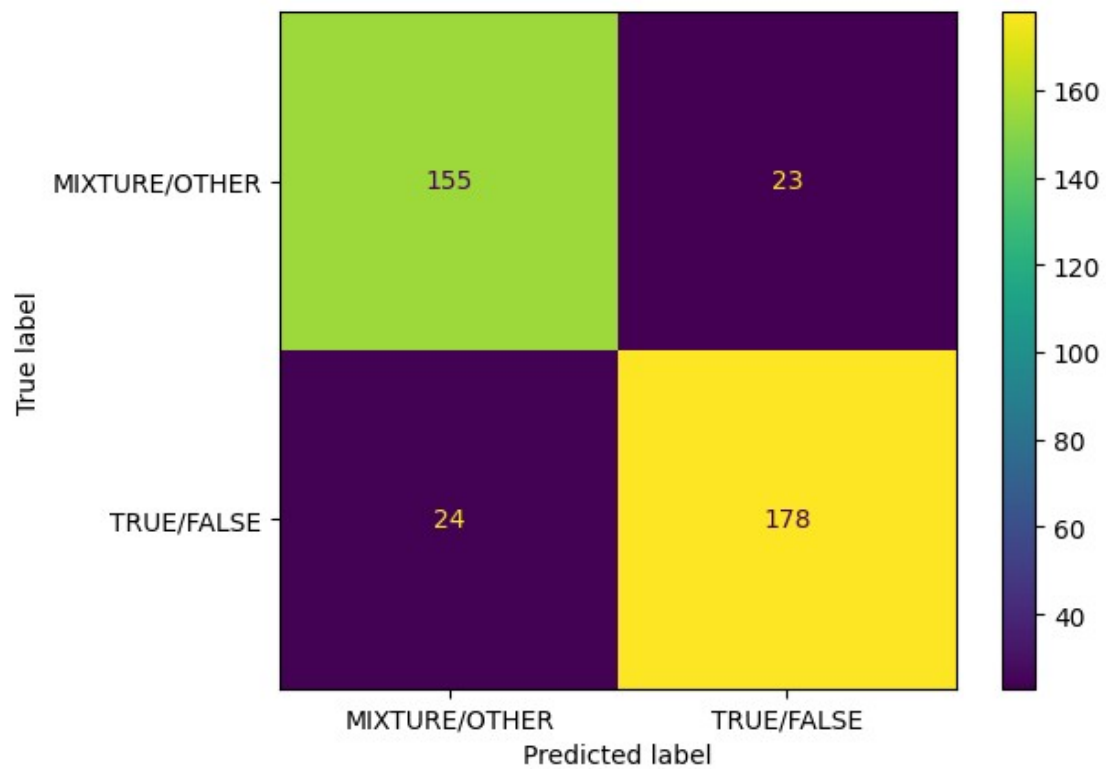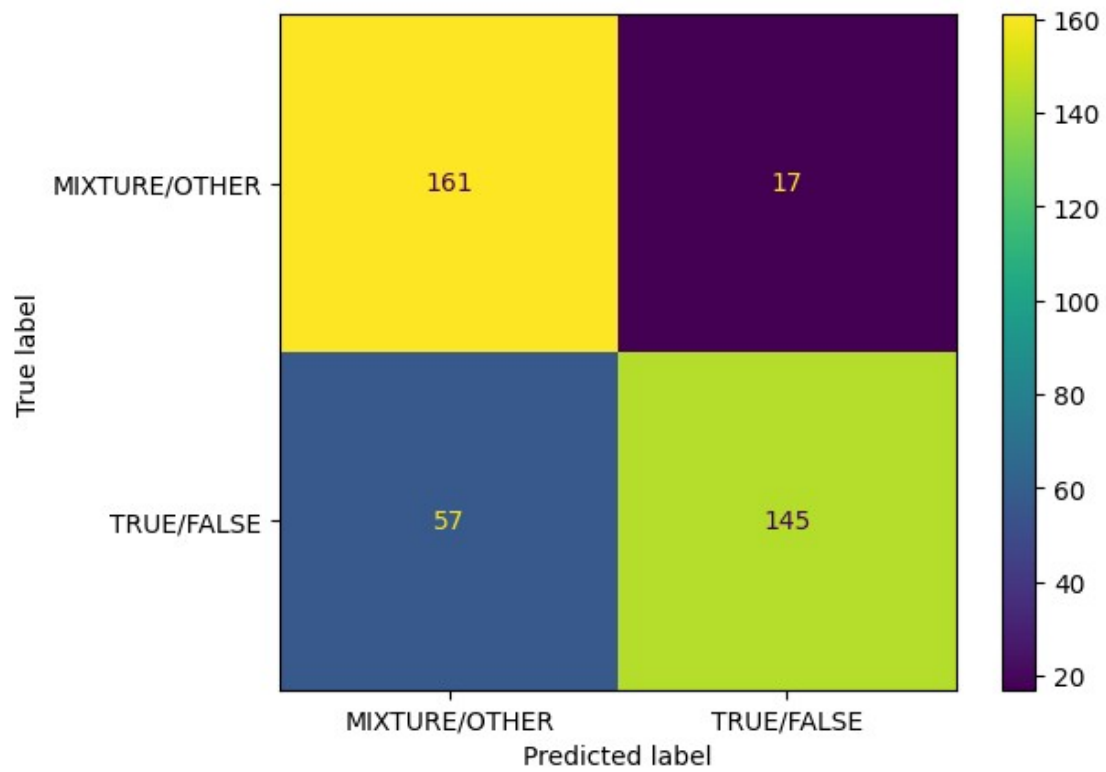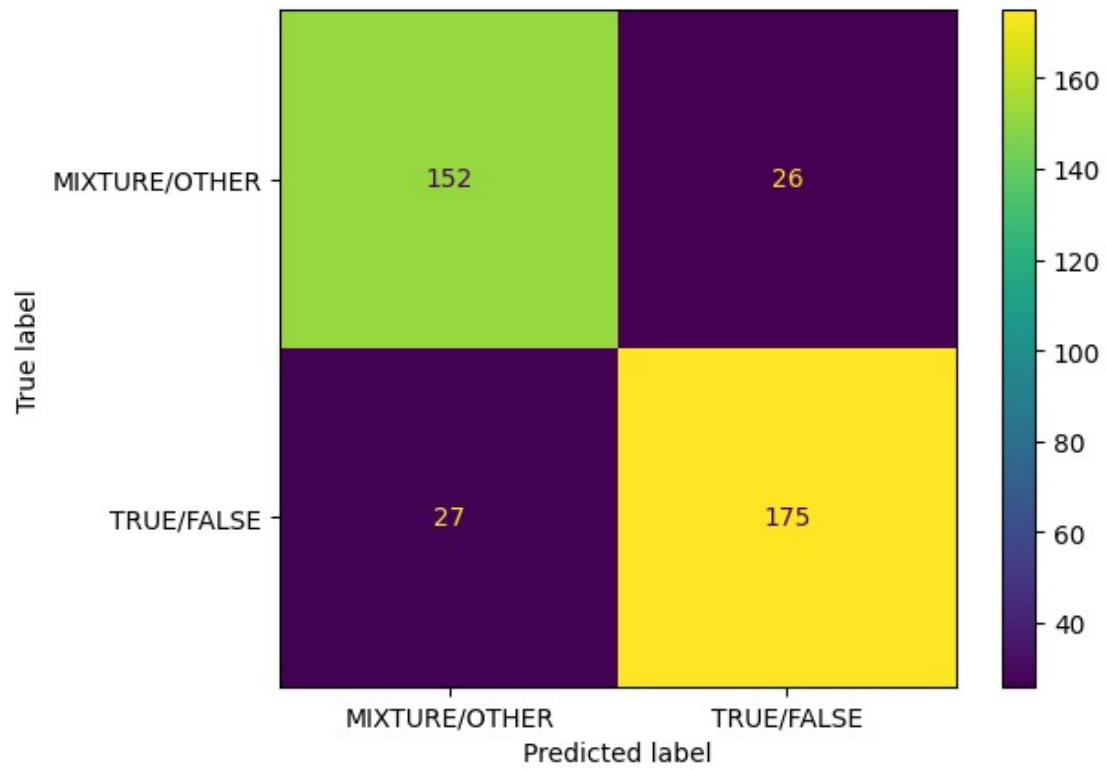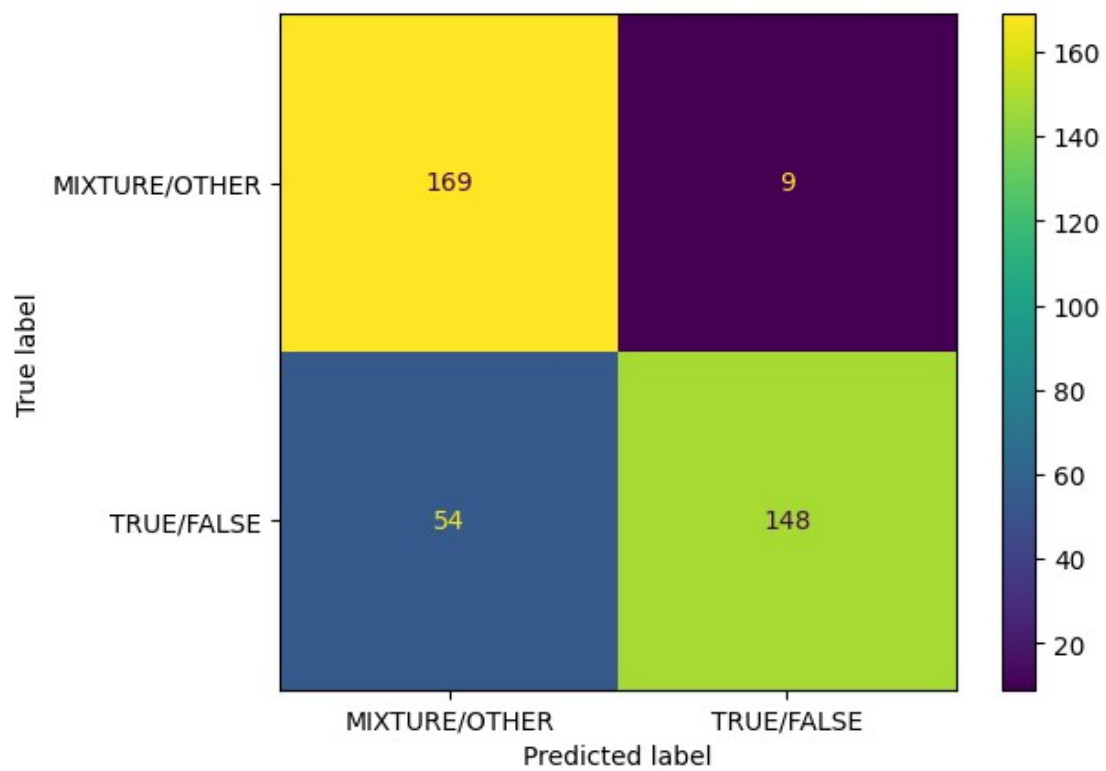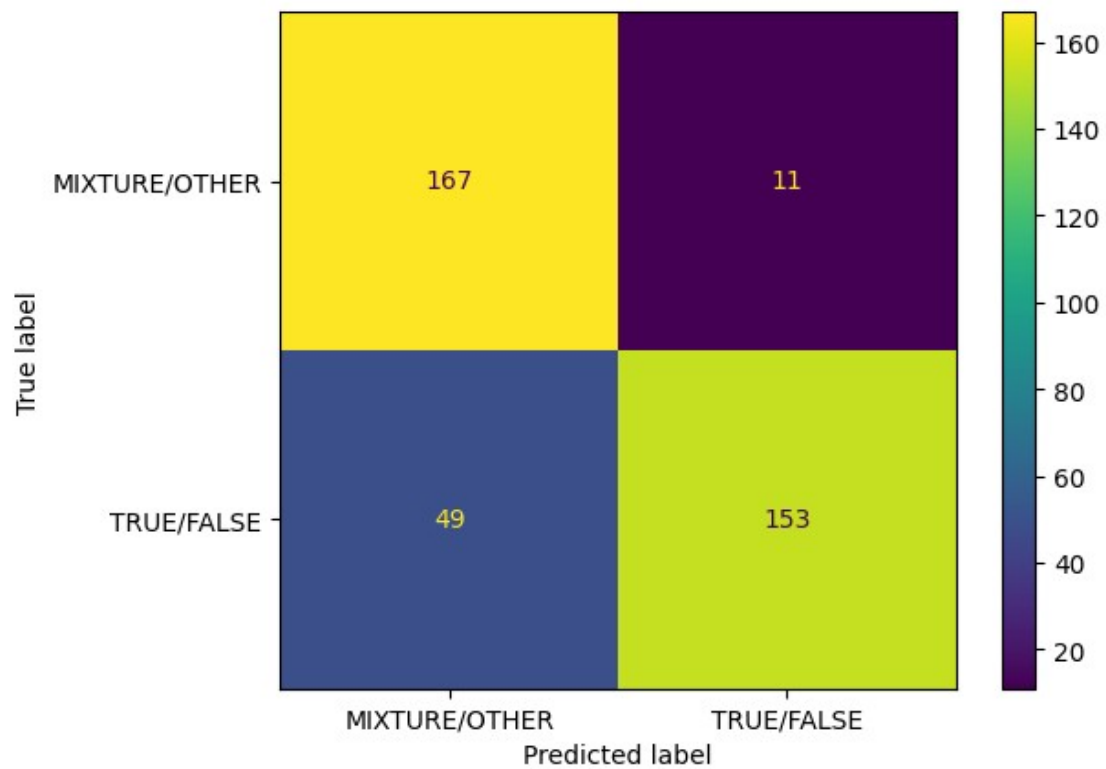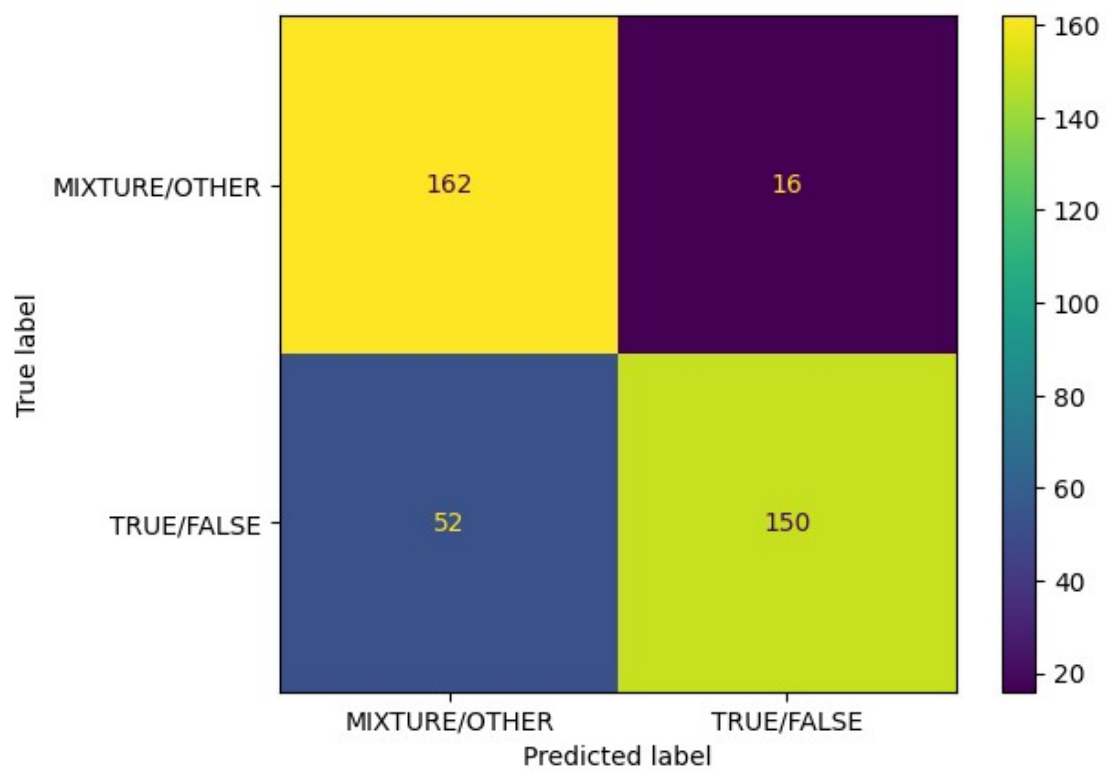
Etape 3 : Classification selon la colonne TEXT

Application d'entités nommées

```python
nlp = spacy.load("en_core_web_sm")
texte=dftrain['text']
#Créer une liste pour stocker les entités nommées
entities = []
# Traiter chaque texte individuellement et ajouter les entités à la
liste
for phrase in texte:
    phrase_str = str(phrase)  # Convertir l'élément en chaîne de
caractères
    doc = nlp(phrase_str)
    for ent in doc.ents:
        entities.append((ent.text, ent.label_))


# Afficher les entités nommées
for entity in entities:
    print(entity[0], entity[1])
```

Le flux de sortie a été tronqué et ne contient que les 5000 dernières
lignes.

Last September DATE
several hundred CARDINAL
John F. Kennedy Memorial Airport PERSON
Ashland GPE
a few miles QUANTITY
GTac GPE
Kennedy PERSON
1963 DATE
One CARDINAL
Mike Wiggins Jr. PERSON
the Bad River LOC
Bad River LOC
Lake Superior LOC
Wiggins PERSON
Walker PERSON
Wisconsin GPE
the State Capitol ORG
Madison GPE
hundreds of thousands CARDINAL
the Bad River LOC
The Bad River LOC
Northern Wisconsin LOC
Last June DATE
one CARDINAL
GTac PERSON
Arizona GPE
Chippewa PERSON
seven CARDINAL
first ORDINAL
35 years DATE
Last summer DATE
Wiggins PERSON
Walker PERSON
Kennedy PERSON
Wiggins PERSON
Kennedy PERSON
Walker PERSON
GTac PERSON
Department of Natural Resources ORG
GTac PERSON
Bill Williams PERSON
Spain GPE
GTac GPE
Williams PERSON
Wiggins PERSON
five CARDINAL
Last August DATE
Obama PERSON
the Interior Department ORG
GTac GPE

five years ago DATE
Obama PERSON
nearly 400 CARDINAL
Native American NORP
this White House FAC
Last week DATE
Wiggins PERSON
two CARDINAL
Helen Harwatt PERSON
Harwatt PERSON
Oregon State University ORG
Bard College ORG
Loma Linda University ORG
American NORP
one CARDINAL
U.S. GPE
2020 DATE
Barack Obama PERSON
2009 DATE
one CARDINAL
between 46 and 74 percent DATE
Harwatt PERSON
Brazilian NORP
Amazon ORG
38,000 CARDINAL
900 metric tons QUANTITY
Brazil GPE
around 212 million CARDINAL
June DATE
U.S. GPE
Brazil GPE
the United Nations ORG
33 percent PERCENT
Earth LOC
26 percent PERCENT
Earth LOC
almost a third CARDINAL
Earth LOC
the next 50 years DATE
the National Academy of Sciences ORG
the United States of America GPE
3 billion CARDINAL
the Associated Press ORG
at least one billion CARDINAL
20 million CARDINAL
Visit Business Insider's ORG
As many as half CARDINAL
2070 DATE
the National Academy of Sciences ORG
the United States of America GPE

the next 50 years DATE
between 1 to 3 billion CARDINAL
2070 DATE
about a third CARDINAL
Marten Scheffer PERSON
Wageningen University ORG
Netherlands GPE
the Associated Press ORG
AP ORG
over 3 billion CARDINAL
at least 1 billion QUANTITY
AP ORG
Cornell University ORG
Natalie Mahowald PERSON
AP ORG
Mahowald PERSON
AP ORG
Scheffer PERSON
AP ORG
1.8 degrees QUANTITY
Fahrenheit WORK_OF_ART
about a billion CARDINAL
annual DATE
the approximately 52 DATE
59 degrees QUANTITY
Fahrenheit WORK_OF_ART
the past 6,000 years DATE
USA Today ORG
the coming 50 years DATE
the past 6,000 years DATE
AP ORG
around 2 billion CARDINAL
the next 50 years DATE
Africa LOC
Asia LOC
South America LOC
Australia GPE
20 million CARDINAL
the Sahara Desert LOC
84 degrees QUANTITY
Fahrenheit WORK_OF_ART
AP ORG
Scheffer PERSON
USA Today ORG
Bernie Sanders PERSON
Vermont GPE
Senate ORG
Democratic NORP
Americans NORP
Fannie ORG

Lou Hamer PERSON
Medgar Evers PERSON
Daisy Bates PERSON
10 to 1 CARDINAL
Americans NORP
Americans NORP
second ORDINAL
Americans NORP
Sixty-five years DATE
Brown v. Board of Education ORG
Topeka GPE
Americans NORP
Princeton PERSON
Keeanga Yamahtta-Taylor PERSON
Black Americans NORP
40 percent PERCENT
2009 DATE
Black Americans NORP
Americans NORP
Americans NORP
15 MONEY
African Americans NORP
42 percent PERCENT
Walton PERSON
40 percent PERCENT
Americans NORP
Joe Biden PERSON
500 CARDINAL
the United States GPE
Rashad Robinson PERSON
Color of Change ORG
America GPE
first ORDINAL
Americans NORP
Walmart ORG
African Americans NORP
one CARDINAL
Coronavirus ORG
three CARDINAL
UK GPE
The UK Addiction Treatment Centres ORG
England GPE
1974 DATE
Wellbutrin NORP
1986 DATE
Melatonin PERSON
Earth LOC
the World Meteorological Organisation ORG
Cedro ORG
Quixadá GPE

Brazil GPE
2016 DATE
the hottest year DATE
2017 DATE
the World Meteorological Organisation ORG
WMO ORG
2016 DATE
Tuesday DATE
El Niño ORG
2016 DATE
US GPE
February DATE
El Niño ORG
2017 DATE
David Carlson PERSON
WMO ORG
Earth LOC
Jeffrey Kargel PERSON
the University of Arizona ORG
US GPE
Prof David Reay PERSON
the University of Edinburgh ORG
Donald Trump PERSON
Trump ORG
Republicans NORP
Congress ORG
Robert Watson PERSON
UK GPE
University of East Anglia ORG
UN ORG
Watson PERSON
Trump ORG
WMO ORG
Petteri Taalas PERSON
1880 DATE
about 115,000 years ago DATE
4m years DATE
2017 DATE
US GPE
February DATE
Australia GPE
Arctic ice LOC
October DATE
six consecutive months DATE
four-decade DATE
Prof Julienne Stroeve PERSON
University College London ORG
UK GPE
Eel PERSON
Lowell's Cove, FAC

Maine GPE
US GPE
Emily Shuckburgh PERSON
the British Antarctic Survey ORG
Arctic LOC
Greenland GPE
Arctic sea LOC
Europe LOC
Asia LOC
North America LOC
between November 2014 and February 2016 DATE
El Niño ORG
15mm QUANTITY
five years DATE
recent decades DATE
2016 DATE
Arctic LOC
tens CARDINAL
Australia GPE
February DATE
Taalas GPE
Sun ORG
roughly 11 years DATE
Sun LOC
2055 DATE
Sun ORG
11 year DATE
NASA ORG
Sun ORG
more than one CARDINAL
seven decades DATE
1645 DATE
1715 DATE
1.3 degrees QUANTITY
Australia GPE
Great Barrier Reef FAC
World Heritage ORG
the Australian Research Council ORG
the Great Barrier Reef EVENT
the mid-2030s DATE
Coral Sea LOC
the Great Barrier Reef EVENT
the past century DATE
March 2016 DATE
Thursday DATE
March 2016 DATE
this year DATE
at least 175 CARDINAL
1 degree Celsius QUANTITY
1.8 degrees QUANTITY

Fahrenheit WORK_OF_ART
March DATE
the Great Barrier Reef FAC
2016 DATE
Catlin Seaview Survey PERSON
year DATE
every two years in the month of March DATE
the year 2034 DATE
the year DATE
2050 DATE
one CARDINAL
the 2030s DATE
the Great Barrier Reef EVENT
Andrew King PERSON
ARC ORG
one CARDINAL
the Great Barrier Reef EVENT
Australians NORP
1992 DATE
U.N. ORG
March 2016 DATE
2034 DATE
Geert Jan van Oldenborgh PERSON
Netherlands GPE
The Paris Agreement ORG
172 CARDINAL
April 22 DATE
American Samoa ORG
first ORDINAL
December 2014 DATE
second ORDINAL
February 2015 DATE
Catlin PERSON
El Niño ORG
third ORDINAL
2014 DATE
2017 DATE
the Coral Sea LOC
Heidi Cullen PERSON
Climate Central ORG
New Jersey GPE
Mashable PERSON
Cullen ORG
the Great Barrier Reef EVENT
the next few decades DATE
March 29 DATE
Australian NORP
the Great Barrier Reef EVENT
The Task Force ORG
10 CARDINAL

Australian NORP
more than 500 CARDINAL
Cairns PERSON
Papua New Guinea GPE
Terry Hughes PERSON
the National Coral Bleaching Taskforce ORG
Hughes ORG
4,000 kilometers QUANTITY
the Great Barrier Reef EVENT
only four CARDINAL
2002 DATE
1998 DATE
one CARDINAL
third ORDINAL
the warmest year DATE
2015 DATE
two CARDINAL
months DATE
Earth LOC
January DATE
February DATE
El Niño ORG
Pacific Ocean LOC
Caucasian Jack PLANO PERSON
TX ORG
seventh ORDINAL
Take Me Out WORK_OF_ART
Ballgame ORG
100 years DATE
caramel ORG
Cracker Jack PERSON
Frito-Lay ORG
today DATE
Caucasian Jack PERSON
Caucasians NORP
the years DATE
Cracker ORG
tens of thousands CARDINAL
Cracker Jacks PERSON
today DATE
Frito-Lay ORG
the last year DATE
the Reserve Bank ORG
8.6 metric tons QUANTITY
first ORDINAL
2009 DATE
New Delhi Sputnik GPE
The Reserve Bank of India RBI ORG
8.46 metric tons QUANTITY
the last fiscal year April 2017 to March 2018 DATE

annual DATE
RBI ORG
566.23 metric tons QUANTITY
June 30, 2018 DATE
557.77 metric tons QUANTITY
June 30, 2017 DATE
first ORDINAL
2009 DATE
200 metric tons QUANTITY
America GPE
India GPE
China GPE
Russia GPE
Belgian NORP
RBI ORG
FER ORG
5.0 percent PERCENT
June 2017 to June 2018 DATE
6.3 percent PERCENT
the previous year DATE
the year DATE
2017-18 DATE
Gold and Ditch Dollar ORG
US GPE
RBI ORG
US GPE
$157 billion MONEY
March DATE
$148.9 billion MONEY
May DATE
RBI ORG
$8.25 billion MONEY
India GPE
13th DATE
US GPE
11th ORDINAL
Greenland GPE
annual DATE
two CARDINAL
Johan Petersen PERSON
Greenland GPE
six metres QUANTITY
annual DATE
Greenland GPE
Greenland GPE
six metres QUANTITY
Greenland GPE
20,000 years ago DATE
GPS ORG
Greenland GPE

12mm QUANTITY
19 cubic kilometres QUANTITY
each year DATE
about 8% PERCENT
Greenland GPE
Greenland GPE
40m years ago DATE
Iceland GPE
Greenland GPE
millennia GPE
Greenland GPE
the past and today DATE
the last 20 years DATE
Greenland GPE
decades DATE
Prof Jonathan Bamber PERSON
the University of Bristol ORG
UK GPE
one CARDINAL
Science Advances ORG
One CARDINAL
Dr Christopher Harig PERSON
the University of Arizona ORG
the end of the century DATE
years ago DATE
Pippa Whitehouse PERSON
the University of Durham ORG
GPS ORG
Greenland GPE
Melt PERSON
Greenland GPE
10 June, 2014 DATE
15 June, 2016 DATE
Every spring DATE
early summer DATE
one CARDINAL
2016 DATE
Antarctica LOC
April DATE
Greenland GPE
August DATE
Ayush PERSON
the Constitution Day DATE
14th April, 2020 DATE
Tulsi PERSON
4 CARDINAL
Dalchini PERSON
2 CARDINAL
Zingiber PERSON
Rhizome 2 PRODUCT

Krishna Marich PERSON
1 CARDINAL
2 CARDINAL
Ayurvedic Pharmacopoeia PERSON
3 grams QUANTITY
500 mg QUANTITY
150 ml QUANTITY
Draksha PERSON
Resins ORG
Lemon Juice ORG
3 CARDINAL
4. CARDINAL
Ayurveda / Siddha /Unani ORG
Drugs & Cosmetics Rules ORG
1945 DATE
States GPE
All State/UT Licensing Authorities and Drug Controllers WORK_OF_ART
ASU Drug Manufacturers ORG
New York City GPE
Chicago GPE
7 month-old DATE
just hours TIME
United Airlines ORG
Monday DATE
about 38 minutes TIME
4th ORDINAL
United Airlines ORG
Pilot WORK_OF_ART
Missouri GPE
United Airlines ORG
38 CARDINAL
LinkedIn Email PERSON
Telegram GPE
Broward County Sheriff PERSON
Scott Israel PERSON
Parkland GPE
Florida GPE
Rick Scott PERSON
next week DATE
Big League Politics ORG
Israel GPE
Florida GPE
Rick Scott PERSON
Big League Politics ORG
IRS ORG
the Sheriff's Office ORG
the last several months DATE
four CARDINAL
Broward County Sheriff ORG
One CARDINAL

Emery Giany PERSON
Republican NORP
Florida GPE
Florida GPE
Pam Bondi PERSON
IG ORG
last week DATE
FBI ORG
Christopher Wray PERSON
Parkland NORP
Wray PERSON
FBI ORG
Parkland GPE
Scott PERSON
Wray PERSON
FBI ORG
Nikolas Cruz PERSON
19 DATE
FBI ORG
Valentine's Day DATE
FBI ORG
FBI ORG
FBI ORG
Scott PERSON
Broward Sheriffs LOC
Israel GPE
FBI ORG
Parkland GPE
Big League Politics ORG
next week DATE
Israel GPE
Parkland NORP
17 CARDINAL
Valentine's Day DATE
Marjory Stoneman Douglas High School ORG
Parkland LOC
45 CARDINAL
the Broward County Sheriff's ORG
FBI ORG
Cruz PERSON
17 CARDINAL
Parkland GPE
FL ORG
Parkland LOC
Democratic NORP
Hillary Clinton PERSON
Debbie Wasserman Schultz PERSON
May DATE
the Broward County Sheriff's ORG
Israel GPE

Brett PERSON
14 year old DATE
Parkland LOC
Cruz PERSON
February 14, 2018 DATE
Israel GPE
the Broward County Sheriff's ORG
Broward ORG
#dodge MONEY
CARDINAL
Republicans NORP
Democrats NORP
the Broward Sheriff's ORG
US GPE
Florida GPE
last week DATE
Florida Agriculture ORG
Adam Putnam PERSON
Ron DeSantis PERSON
Parkland NORP
Putnam ORG
DeSantis PERSON
Scott Israel PERSON
the Second Amendment LAW
Broward County GPE
Putnam ORG
Scott PERSON
the coming week DATE
Sheriff Israel PERSON
Scott PERSON
US Senate ORG
Florida GPE
Senate ORG
the days DATE
Parkland LOC
Scott Israel PERSON
the Democratic Party ORG
FBI ORG
Parkland GPE
Sheriff Israel PERSON
FBI ORG
NHS ORG
England GPE
more than £1 billion MONEY
NHS Digital ORG
more than £422 million MONEY
the last 10 years DATE
Almost one CARDINAL
20 CARDINAL
2 CARDINAL

around 90% PERCENT
Robin Hewings PERSON
UK GPE
the last 20 years DATE
26,000 CARDINAL
1bn ORDINAL
NHS ORG
Hewings PERSON
2 CARDINAL
years earlier' DATE
Postcode PERSON
Presentational NORP
over three million CARDINAL
England GPE
the last two decades DATE
nearly 100,000 CARDINAL
92% PERCENT
2 CARDINAL
2 CARDINAL
1 CARDINAL
Almost seven CARDINAL
10 CARDINAL
66.8% PERCENT
almost six CARDINAL
10 CARDINAL
57.8% PERCENT
England GPE
2 CARDINAL
more than 80 percent PERCENT
obese NORP
2 CARDINAL
NHS England LOC
Nearly £477 million MONEY
2017-18 DATE
the same year DATE
around £350 million MONEY
£181 million MONEY
Jonathan Valabhji PERSON
NHS England FAC
NHS ORG
2 CARDINAL
first ORDINAL
The NHS Diabetes Prevention Programme ORG
million CARDINAL
2 CARDINAL
Michigan GPE
the Great Lakes Justice Center FAC
two CARDINAL
Detroit GPE
Wayne County GPE

Michigan GPE
David A. Kallman PERSON
Court ORG
Wayne County GPE
Michigan GPE
Pennsylvania GPE
two CARDINAL
Donald Trump PERSON
Naomi Wolf PERSON
Cheryl A. Costantino PERSON
Edward P. McCall Jr. PERSON
the Qualified Voter File QVF PERSON
November 3 DATE
2020 CARDINAL
Michigan GPE
The Western Journal ORG
Privacy Policy and Terms of Use You ORG
100% PERCENT
12 CARDINAL
Votes ORG
0% PERCENT
the TCF Center FAC
tens of thousands CARDINAL
Democratic NORP
the TCF Center FAC
Joe Biden PERSON
Federal Investigators Arrive ORG
Michigan GPE
the TCF Center FAC
tens of thousands CARDINAL
thousands CARDINAL
Detroit GPE
two CARDINAL
David Fink PERSON
The Detroit News ORG
last week DATE
The Western Journal ORG
50 CARDINAL
Torbay PERSON
Devon PERSON
51 per cent MONEY
North Yorkshire GPE
51 per cent MONEY
Redbridge PERSON
London GPE
less than 1 CARDINAL
Southampton GPE
3 per cent MONEY
North Yorkshire GPE
51 per cent MONEY

BBC ORG
yesterday DATE
35 CARDINAL
Isle of Wight LOC
Jon Platt PERSON
the High Court ORG
108 CARDINAL
35 CARDINAL
28 CARDINAL
five CARDINAL
David Hedley PERSON
2014-15 DATE
Shropshire GPE
Herefordshire PERSON
Bracknell Forest ORG
Leicester GPE
Barnsley ORG
Luton GPE
Rochdale and Bexley ORG
London GPE
season DATE
thousands CARDINAL
2013 DATE
Michael Gove PERSON
60 MONEY
2014-15 DATE
May 2016 DATE
Isle of Wight LOC
Jon Platt PERSON
the High Court ORG
the year DATE
2014-15 DATE
3.2million school days DATE
two years DATE
4.1million CARDINAL
just under 900,000 CARDINAL
The Department for Education ORG
the High Court's ORG
the Supreme Court ORG
later this year DATE
A Department for Education ORG
every extra day DATE
GCSEs ORG
recent years DATE
Redbridge Council ORG
GCSE ORG
One CARDINAL
Darren Paffey PERSON
Southampton Council ORG
Southampton GPE

first ORDINAL
Sea LOC
Elizabeth Keatinge PERSON
Earth LOC
Arctic LOC
Antarctic LOC
January DATE
this week DATE
the summer DATE
winter DATE
Arctic sea LOC
this January DATE
5.17 million square miles QUANTITY
the month DATE
38-year DATE
the National Snow and Ice Data Center ORG
100,000 square miles QUANTITY
the previous January DATE
just last year DATE
Arctic sea LOC
January 2017 DATE
January DATE
38 CARDINAL
NSIDC News (@NSIDC ORG
February 7, 2017 January DATE
the Arctic Ocean LOC
NASA ORG
Arctic LOC
9 degrees QUANTITY
the month DATE
Antarctica LOC
Antarctic LOC
summer DATE
the Amundsen Sea LOC
Arctic LOC
U.S. GPE
Sea LOC
the Southern Hemisphere LOC
January 2017 DATE
#Antarctica https://t.co/3pPss4vRVJ pic.twitter.com/IKvLna3Ull MONEY
February 7, 2017 DATE
summer DATE
Arctic LOC
the past few decades DATE
the National Oceanic and Atmospheric Administration ORG
the latter half of the 20th century DATE
Antarctic LOC
wildly year to year DATE
NASA ORG
Walt Meier PERSON

thousands of years DATE
BUFFALO GPE
N.Y. GPE
New York State GPE
New York GPE
one CARDINAL
only 12 CARDINAL
gallon CARDINAL
more than $2.00 MONEY
Nationally PERSON
gallon CARDINAL
1.77 MONEY
2.89 MONEY
a year ago DATE
New York State GPE
2.18 MONEY
2.96 MONEY
a year ago DATE
Buffalo GPE
2.23 MONEY
gallon CARDINAL
2.88 MONEY
a year ago DATE
Elizabeth Carey PERSON
AAA ORG
China GPE
nine straight weeks DATE
Carey PERSON
Western New Yorkers NORP
second ORDINAL
NYS ORG
97 cents MONEY
46 cents MONEY
1.43 MONEY
gallon CARDINAL
New York GPE
Carey PERSON
the Empire State GPE
New Yorkers NORP
0.65 MONEY
more than $0.45 MONEY
Ohio GPE
about half CARDINAL
New Yorkers NORP
Robin Vos PERSON
the Wisconsin State Assembly ORG
Vos PERSON
Wisconsin GPE
almost half CARDINAL
99 CARDINAL

November 2014 DATE
half CARDINAL
August DATE
the 12th DATE
Assembly District LOC
9 CARDINAL
134 CARDINAL
Minnesota GPE
Michigan GPE
110 CARDINAL
State ORG
Wisconsin GPE
State Assembly ORG
Soviet NORP
Politburo ORG
House of Lords ORG
American NORP
Wisconsin GPE
the Government Accountability Board ORG
tens of thousands CARDINAL
2016 DATE
half CARDINAL
Wisconsin GPE
Wisconsin GPE
Citizens ORG
2014 DATE
Representatives ORG
Wisconsin GPE
Wisconsin GPE
a State Assembly ORG
Representatives ORG
Andy Craig PERSON
Libertarian Party of Wisconsin ORG
Midwest Regional LOC
Competitive Elections Wisconsin ORG
the Libertarian Party Competitive Elections Wisconsin ORG
the Libertarian Party of Wisconsin ORG
Our America Initiative LOC
Libertarian Congressional ORG
IPR ORG
Andy Craig PERSON
GETTY ORG
Thousands CARDINAL
SUBSCRIBE Invalid PRODUCT
Network Janet PERSON
dozens CARDINAL
UK GPE
Daesh GPE
Monday DATE
morning TIME

Distributed Denial of Service ORG
thousands CARDINAL
Manchester PERSON
Edinburgh PERSON
Sussex GPE
Glasgow PERSON
Royal Holloway PERSON
Network Janet PERSON
WiFi PERSON
tens of thousands CARDINAL
GETTY Students ORG
hours TIME
The Tab PERSON
the University of Manchester IT ORG
Janet PERSON
this morning TIME
Jisc PERSON
ISIS ORG
Royal Holloway PERSON
UCL ORG
University College London UCL ORG
yesterday DATE
WiFi PERSON
UCL ORG
GETTY Students ORG
ISIS ORG
Second year DATE
Emma PERSON
Edinburgh GPE
Manchester third year DATE
Tom PERSON
Ali G PERSON
WiFi PERSON
third year DATE
Claire PERSON
ISIS ORG
Fri, March 25, 2016 DATE
Iraq GPE
Syrian NORP
Paris GPE
Brussels GPE
1 CARDINAL
47 CARDINAL
Belgian NORP
Salah Abdeslam PERSON
Paris GPE
WASHINGTON GPE
The House of Representatives ORG
today DATE
217-212 CARDINAL

H.R. 1628 DATE
the American Health Care Act ORG
Republican NORP
the Affordable Care Act LAW
the Congressional Budget Office ORG
CBO ORG
24 million CARDINAL
Americans NORP
the next decade DATE
seven million CARDINAL
Americans NORP
$838 billion MONEY
Medicaid ORG
Americans NORP
up to five CARDINAL
Planned Parenthood ORG
one year DATE
Americans NORP
$1 trillion MONEY
today DATE
annual DATE
Chu PERSON
Republicans NORP
millions CARDINAL
the Congressional Budget Office ORG
the White House ORG
Trump ORG
California GPE
One CARDINAL
California GPE
Californians NORP
7 million CARDINAL
Medicare ORG
50-64 DATE
5 CARDINAL
24 million CARDINAL
Americans NORP
nearly 2 million CARDINAL
Asian NORP
Americans NORP
Pacific Islanders NORP
Planned Parenthood ORG
one year DATE
Trumpcare ORG
$838 billion MONEY
Medicaid ORG
Trumpcare ORG
400 CARDINAL
America GPE
$2.8 million MONEY

Democrats NORP
Affordable Care Act LAW
Americans NORP
Americans NORP
Cassandra Fairbanks PERSON
Jim Hoft PERSON
Tuesday, November 3rd DATE
Trump PERSON
Joe Biden PERSON
Michigan GPE
over 100,000 CARDINAL
Trump PERSON
Michigan GPE
2016 DATE
Detroit GPE
Michigan GPE
election night TIME
the TCF Center FAC
Cobo Hall FAC
Detroit City GPE
GOP ORG
GOP ORG
hundreds CARDINAL
At least three CARDINAL
the TCF Center ORG
early in the morning TIME
November 4th DATE
PUNK PERSON
Dan Bongino PERSON
Michigander Shane Trejo PERSON
3:30 am TIME
Biden Ballot PERSON
Joe Biden PERSON
thousands CARDINAL
Trejo PERSON
at least 50 CARDINAL
3:30 am TIME
8:00 pm TIME
3:30 am TIME
first ORDINAL
10:30 pm TIME
138,000 CARDINAL
Detroit Clerk's Office ORG
Janet Winfrey's PERSON
Trejo PERSON
the day DATE
Joe Biden PERSON
Michigan GPE
Shane Trejo PERSON
The Biden Ballot Drop ORG

one CARDINAL
Trejo PERSON
Michigan GPE
Pat Colbeck PERSON
the TCF Center FAC
the morning TIME
November 4th DATE
Colbeck PERSON
3:30 am TIME
Biden Ballot Drop PERSON
The Gateway Pundit WORK_OF_ART
November DATE
Detroit Election Bureau ORG
Republican NORP
Democrat NORP
Coney Island LOC
Michigan GPE
today DATE
the TCF Center ORG
election night TIME
The Gateway Pundit ORG
TCF ORG
December DATE
The TCF Center ORG
over $22,000 MONEY
one day DATE
two hours TIME
Last Friday DATE
last week DATE
the hundreds of hours TIME
the TCF Center ORG
the morning TIME
November 4th DATE
3:23 AM TIME
The Gateway Pundit ORG
Detroit GPE
the TCF Center FAC
3:25 AM TIME
second ORDINAL
three CARDINAL
over 50 CARDINAL
the TCF Building FAC
about 25 minutes later TIME
an hour later TIME
the TCF Center ORG
TCF ORG
Detroit GPE
Michigan GPE
GOP ORG
November DATE

the TCF Center ORG
the coming days DATE
the Detroit TCF Center ORG
election night TIME
Cassandra Fairbanks ORG
Judicial Watch ORG
Detail Role PERSON
Justice Department ORG
Trayvon Martin PERSON
Sanford PERSON
FL ORG
17-year-old DATE
African American NORP
Washington, DC GPE
Judicial Watch ORG
today DATE
the Department of Justice ORG
the Community Relations Service ORG
Sanford PERSON
FL ORG
Trayvon Martin PERSON
George Zimmerman PERSON
weekly DATE
AL AK PERSON
CA CO CT DE DC FL ORG
IL ORG
IA GPE
LA ME ORG
MA MI ORG
MN MS ORG
MT NE NV NM NH NJ ORG
NC GPE
VT VA WA WV FAC
WI WY Judicial ORG
Weekly Update JW ORG
Freedom of Information Act ORG
FOIA ORG
April 24, 2012 DATE
125 CARDINAL
May 30, 2012 DATE
JW ORG
June 5, 2012 DATE
222 CARDINAL
March 6, 2013 DATE
March 25 DATE
2012 DATE
CRS ORG
674.14 MONEY
Sanford PERSON
FL ORG

African-American NORP
March 25 DATE
2012 DATE
CRS ORG
1,142.84 MONEY
Sanford GPE
FL ORG
African-American NORP
March 30 DATE
April 1, 2012 DATE
CRS ORG
892.55 MONEY
Sanford GPE
FL ORG
Florida GPE
three CARDINAL
Florida GPE
March 30 DATE
April 1, 2012 DATE
CRS ORG
an additional $ MONEY
751.60 MONEY
Sanford GPE
FL ORG
the City of Sanford GPE
March 31 DATE
April 3 — 12, 2012 DATE
CRS ORG
1,307.40 MONEY
Sanford GPE
FL ORG
Sanford PERSON
April 11 DATE
12, 2012 DATE
CRS ORG
552.35 MONEY
Sanford GPE
FL ORG
17 year DATE
African American NORP
Florida GPE
Sunshine Law PERSON
April 23, 2012 DATE
JW ORG
thousands CARDINAL
April 27, 2012 DATE
Miami-Dade County Community Relations Board Program GPE
Amy Carswell PERSON
April 16, 2012: DATE
Thomas Battles PERSON

Mildred De Robles PERSON
Miami-Dade Coordinator GPE
the U.S. Department of Justice Community Relations Service ORG
Sanford GPE
Florida GPE
the Orlando Sentinel ORG
Battles ORG
Miami GPE
Tommy PERSON
the Civil Rights Act LAW
1964 DATE
Barack Obama PERSON
Judicial Watch ORG
Sanford PERSON
FL ORG
Zimmerman PERSON
April 15, 2012 DATE
the Orlando Sentinel ORG
NAACP ORG
Bill Lee PERSON
Turner Clayton ORG
Seminole County GPE
the National Association for the Advancement of Colored People ORG
Valarie Houston PERSON
Allen Chapel AME Church ORG
March 20 DATE
Florida GPE
Sunshine Law PERSON
the City of Sanford GPE
Judicial Watch ORG
Second Shiloh Missionary Baptist Church ORG
Sanford PERSON
April 19, 2012 DATE
Sanford PERSON
Bill Lee PERSON
the "Dream Defenders ORG
Lee PERSON
the Orlando Sentinel ORG
CRS ORG
40-mile QUANTITY
Daytona Beach GPE
Sanford PERSON
the Justice Department ORG
George Zimmerman PERSON
Judicial Watch ORG
Tom Fitton PERSON
Americans NORP
the Great Divide FAC
nine CARDINAL
nine CARDINAL

American NORP
Florida GPE
post-Obama NORP
Minnesota GPE
Wisconsin GPE
Latino PERSON
Texas GPE
Nevada GPE
GOP ORG
Washington GPE
Republican NORP
South LOC
Kentucky GPE
Colorado GPE
one CARDINAL
Democrats NORP
Nebraska GPE
Louis Alvarez PERSON
Andrew Kolker PERSON
Paul Stekler PERSON
six CARDINAL
PBS ORG
Election 2016 DATE
The Washington Post ORG
the Corporation for Public Broadcasting ORG
Latino Public Broadcasting ORG
PBS ORG
the World Channel ORG
Abnormal PERSON
New Orleans GPE
Katrina EVENT
Stacy Head PERSON
city council ORG
Katrina EVENT
African-American NORP
Barbara Lacen-Keller PERSON
New Orleans GPE
New-Orleans NORP
Abnormal PERSON
America GPE
Louis Alvarez PERSON
Andy Kolker PERSON
Peter Odabashian PERSON
America GPE
Louisiana GPE
South LOC
Southwest LOC
2013 DATE
PBS ORG
POV GPE

2014 DATE
The Los Angeles Times ORG
the New Orleans Advocate ORG
The Boston Globe ORG
New Orleans GPE
Austin American-Statesman PERSON
The Washington Post ORG
fourth ORDINAL
Stekler PERSON
POV GPE
one CARDINAL
the last four decades DATE
Stekler PERSON
New Orleans GPE
the Crescent City GPE
Abnormal PERSON
The International Documentary Association's ORG
New Orleans GPE
Post-Katrina ORG
Abnormal' Looks at Life in the Crescent City WORK_OF_ART
' 2014 DATE
POV GPE
Choice 2008 DATE
two-hour TIME
PBS ORG
Obama-McCain ORG
Michael Kirk PERSON
Jim Gilmore PERSON
Barack Obama LOC
John McCain PERSON
three CARDINAL
October 2008 DATE
an estimated 7.8 million MONEY
an additional 3.5 million QUANTITY
The St. Louis Post-Dispatchdescribed ORG
The Boston Globe ORG
these two hours TIME
the Baltimore Sun ORG
two-hour TIME
30 minutes TIME
the Los Angeles Times ORG
PBS ORG
Stekler PERSON
PBS ORG
P.O.V. ORG
their 2004 season DATE
Sandra Guardado PERSON
Texas GPE
George W. Bush PERSON
the White House ORG

2002 DATE
Lyndon Johnson's PERSON
Bush PERSON
Lone Star ORG
Republican Party ORG
Democratic NORP
State Rep ORG
Rick Green PERSON
Patrick Rose PERSON
Karl Rove PERSON
Ann Richards PERSON
Molly Ivins PERSON
two CARDINAL
Rick Green PERSON
Patrick Rose PERSON
late on election night TIME
one CARDINAL
Liz Smith PERSON
USA GPE
Variety ORG
Texas GPE
IndieWIRE ORG
about two CARDINAL
Texas GPE
The Washington Post ORG
the Dallas Morning News ORG
John Leonard PERSON
New York Magazine GPE
second ORDINAL
De Tocqueville PERSON
Paul Stekler PERSON
2004 DATE
POV GPE
George Wallace PERSON
Settin PERSON
the Woods on Fire ORG
Daniel McCabe PERSON
2000 DATE
Sundance Film Festival EVENT
2001 DATE
three-hour TIME
two-night TIME
PBS ORG
The American Experience WORK_OF_ART
April 23-24, 2000 DATE
Wallace PERSON
Lurleen PERSON
The New York Times ORG
the Houston Chronicle ORG
the Memphis Commercial Appeal ORG

the National Journal ORG
The Boston Globe ORG
the Toronto Star ORG
The Wall Street Journal ORG
'The Sopranos' WORK_OF_ART
the Texas Observer ORG
George Wallace PERSON
Settin PERSON
Five CARDINAL
Providence GPE
Buddy Cianci ORG
America GPE
Louis Alvarez PERSON
Andrew Kolker PERSON
two-night TIME
four-hour TIME
American NORP
thirty CARDINAL
the White House ORG
WETA GPE
the Los Angeles Times ORG
the Chicago Tribune ORG
Americana PERSON
USA Today ORG
four CARDINAL
Roll Call PERSON
the Atlanta Journal Constitution ORG
Vote for Me WORK_OF_ART
1 CARDINAL
2 CARDINAL
Stekler PERSON
Louis Alvarez PERSON
Andy Kolker PERSON
Louisiana GPE
hour TIME
Byzantine NORP
Louisiana GPE
the days DATE
Alfred I. duPont-Columbia PERSON
Journalism Award PERSON
Ohio State Award ORG
Chicago International Film Festival Silver Plaque ORG
National Educational Film Festival Silver Apple ORG
Emmy PERSON
PBS ORG
P.O.V. ORG
August 1992 DATE
Edwin PERSON
two CARDINAL
1992 DATE

1 CARDINAL
2 CARDINAL
weekly DATE
PBS ORG
Texas GPE
biannual DATE
MTV ORG
The Real World ORG
Austin PERSON
season DATE
SXSW ORG
Paul Stekler PERSON
Diane Zander-Mason PERSON
Megan Field GPE
Sandra Guardado PERSON
Megan Field ORG
Sandra Guardado PERSON
Deborah Eve Lewis PERSON
Producer GPE
Paul Stekler PERSON
Native-American NORP
James Welch PERSON
Winter DATE
the Blood, Fool's Crow GPE
this hour TIME
Custer PERSON
Last Stand PERSON
Big Horn LOC
Montana GPE
Native-American NORP
PBS ORG
THE AMERICAN EXPERIENCE ORG
November 1992 DATE
that year DATE
two CARDINAL
the Spur Award WORK_OF_ART
the Western Writer's of America ORG
the Little Big Horn National Monument ORG
James Welch's PERSON
Killing Custer: The Battle of the Little Bighorn and the Fate of the
Plains Indians WORK_OF_ART
Purchase Film James Welch WORK_OF_ART
1940-2003 DATE
Paul Stekler(The Austin Chronicle ORG
August 22, 2003 DATE
Custer PERSON
Producer GPE
Writer PERSON
Jacqueline Shearer PERSON
two hour TIME

Blackside ORG
The Promised Land: 1967-68 WORK_OF_ART
Martin Luther King's PERSON
last year DATE
two CARDINAL
CINE Golden Eagle WORK_OF_ART
Banff Television Festival ORG
The Keys to the Kingdom: 1972-78 WORK_OF_ART
Boston GPE
Atlanta GPE
first ORDINAL
Series EVENT
the Organization of American Historians' ORG
Eric Barnouw Award PERSON
PBS ORG
January 1990 DATE
Purchase Film ORG
Henry Hampton PERSON
1940-1998 DATE
Memory WORK_OF_ART
Paul Stekler PERSON
The Austin Chronicle ORG
January 1, 1999 DATE
PW: Memphis ORG
Producer GPE
hour TIME
New Orleans GPE
two CARDINAL
Winner PERSON
American Film Festival Red Ribbon EVENT
National Educational Film Festival Bronze Apple ORG
PBS ORG
September 1987 DATE
Producer GPE
Writer PERSON
Editor PERSON
Alan Bell PERSON
hour TIME
the Mississippi Delta LOC
Winner PERSON
a Gabriel Award PERSON
the American Film Festival EVENT
PBS ORG
February 1985 DATE
SHORTS PERSON
the last few years DATE
Stekler PERSON
Love Me WORK_OF_ART
Margaret Brown PERSON
Texas GPE

Townes Van Zandt PERSON
the Toronto Film Festival EVENT
Peter Frumkin PERSON
PBS ORG
American Masters WORK_OF_ART
One CARDINAL
Spit Farther PERSON
annually DATE
Luling GPE
Texas GPE
2002 DATE
Building Futures ORG
SCALEAfrica at Chiutika Basic School ORG
Mfuwe GPE
Zambia GPE
two CARDINAL
Beauty Banda PERSON
three CARDINAL
Building Futures ORG
one CARDINAL
UT Austin's FAC
Blanton Museum of Art ORG
fifty-year DATE
the Texas Observer ORG
Ebeneezer ORG
the Amuzu Hospital FAC
at least 21 CARDINAL
at least 21 CARDINAL
Harvard University ORG
the European Urology ORG
Ted Cruz PERSON
Texas GPE
the United States GPE
Senate ORG
CNN ORG
the G-7 Summit ORG
Italy GPE
Trump PERSON
the Paris Agreement ORG
United Nations ORG
Trump PERSON
Obama GPE
last year DATE
the United States GPE
National Economic Research Associates Economic Consulting ORG
the Paris Agreement ORG
$3 trillion MONEY
6.5 million CARDINAL
7,000 MONEY
American NORP

2040 DATE
2025 DATE
$250 billion MONEY
2.7 million CARDINAL
21% PERCENT
19% PERCENT
11% PERCENT
American NORP
monthly DATE
the United States GPE
American NORP
2016 DATE
American NORP
first ORDINAL
82% PERCENT
the United States GPE
almost one-third CARDINAL
America GPE
America GPE
the next decade DATE
The Clean Power Plan LAW
Texans NORP
16% PERCENT
2030 DATE
the Economic Reliability Council of Texas ORG
thousands CARDINAL
Americans NORP
the Paris Agreement ORG
EPA ORG
America GPE
less than two-tenths CARDINAL
America GPE
Obama GPE
America GPE
Russia GPE
approximately 50% PERCENT
China GPE
India GPE
2030 DATE
American NORP
The Paris Agreement ORG
America GPE
Russia GPE
Arctic LOC
China GPE
Arctic LOC
American NORP
America GPE
US GPE
today DATE

Twitter PRODUCT
Facebook Efforts ORG
US GPE
the Paris Agreement ORG
Trump PERSON
America GPE
the Paris Agreement ORG
Joe Biden PERSON
more than 2,616 CARDINAL
BLM ORG
Biden PERSON
Biden PERSON
the White House ORG
BLM ORG
Art Tubolls PERSON
Trump ORG
morning TIME
Circle K PERSON
2 MONEY
Circle K PERSON
Joe Biden PERSON
Donald Trump PERSON
America GPE
Biden PERSON
One CARDINAL
today DATE
Biden PERSON
nearly 3000 CARDINAL
a long four years DATE
America GPE
Dominic Raab PERSON
Muhammad Ali PERSON
Mother Teresa PERSON
20% PERCENT
the past 25 years DATE
Britain GPE
post-Brexit NORP
The Sunday Times ORG
Dominic Raab PERSON
the Migration Advisory Committee ORG
MAC ORG
this autumn DATE
Raab PERSON
Brexiteer PERSON
Tory NORP
first ORDINAL
Tory NORP
November 24, 2013 DATE
Theodore Roosevelt PERSON
1858-1919 DATE

about 2007 DATE
Roosevelt PERSON
half CARDINAL
at least 1995 DATE
Google Groups ORG
BigBird/Barney SUCKS-PBSgate William Davenant 2/9/95 ORG
Brown Cafe UPS Forum Raw ORG
08:28 TIME
ANGER ORG
TRUTH ORG
TED PERSON
November 18, 2008 DATE
12:42 AM TIME
steve orris PERSON
@steveorris PRODUCT
3:15 PM - 3 TIME
08 CARDINAL
7/23/2009 DATE
12:26:45 PM TIME
Canada GPE
Teddy Roosevelt PERSON
Franc Pohole ORG
Conservative GPE
Liberal GPE
8:21 PM - 19 TIME
Sep ORG
07:19 PM TIME
Teddy Roosevelt PERSON
Theodore Roosevelt PERSON
today DATE
LIE ORG
Teddy Roosevelt PERSON
6:00 PM - 1 Nov 13 TIME
£10 billion MONEY
Duncan Smith PERSON
yesterday DATE
Universal Support ORG
Iain Duncan Smith PERSON
10.1billion MONEY
Universal Credit ORG
Duncan Smith PERSON
Through Universal Support WORK_OF_ART
the Lankelly Chase Foundation ORG
£10.1 billion MONEY
Duncan Smith PERSON
Church ORG
586,000 CARDINAL
at least one CARDINAL
Duncan Smith PERSON
first ORDINAL

the Department of Health ORG
60 per cent MONEY
Duncan Smith PERSON
Church ORG
478,000 CARDINAL
2015 DATE
almost 700,000 CARDINAL
2010 DATE
4:22 P.M. EST TIME
Good afternoon TIME
two CARDINAL
today DATE
COVID-19 ORG
a few days ago DATE
the Food and Drug Administration ORG
Johnson & Johnson COVID-19 ORG
third ORDINAL
COVID-19 PERSON
one CARDINAL
first ORDINAL
100 million CARDINAL
my first 100 days DATE
37 days DATE
Johnson & Johnson ORG
two CARDINAL
today DATE
Two CARDINAL
Johnson & Johnson ORG
Merck ORG
Johnson & Johnson's ORG
World War Two EVENT
the Defense Production Act LAW
two CARDINAL
Merck ORG
J&J NORP
Johnson & Johnson ORG
Johnson & Johnson's ORG
24/7 CARDINAL
the Defense Production Act LAW
the Department of Defense ORG
Johnson & Johnson's ORG
Johnson & Johnson ORG
Merck ORG
America GPE
the end of May DATE
America GPE
About three weeks ago DATE
the end of July DATE
today DATE
America GPE

the end of May DATE
the end of May DATE
millions CARDINAL
Americans NORP
more than 1,500 CARDINAL
the Federal Emergency Management Agency ORG
FEMA ORG
the Commissioned Health Corps ORG
the Department of Health and Human Services ORG
the Defense Department ORG
the National Guard ORG
thousands CARDINAL
millions CARDINAL
over 7,000 CARDINAL
COVID-19 ORG
hundreds CARDINAL
thousands CARDINAL
Jill PERSON
Houston GPE
last week DATE
three CARDINAL
year DATE
millions CARDINAL
each day DATE
December DATE
American NORP
Congress ORG
House ORG
American NORP
last week DATE
Senate ORG
the Centers for Disease Control and Prevention ORG
CDC ORG
Congress ORG
today DATE
one CARDINAL
's GPE
today DATE
the District of Columbia GPE
at least one CARDINAL
the end of the month of March DATE
next week DATE
a month DATE
the month of March DATE
12 CARDINAL
March DATE
March DATE
the first week DATE
this month DATE
the month of March DATE

today DATE
recent winter DATE
Midwest LOC
South LOC
third ORDINAL
Americans NORP
my first 100 days DATE
today DATE
more than 511,839 CARDINAL
Americans NORP
today DATE
the United States GPE
Senate ORG
House ORG
American NORP
American NORP
FEMA ORG
American NORP
God PERSON
next year DATE
today DATE
4:34 P.M. EST TIME
Turku University ORG
Finland GPE
0.1 CARDINAL
the last century DATE
just 0.01 CARDINAL
Jyrki Kauppinen PERSON
the last hundred years DATE
about 0.1 CARDINAL
about 0.01 CARDINAL
Kauppinen PERSON
Malmi GPE
UN ORG
dozens CARDINAL
the United Nations ORG
YouTube ORG
SubscribeStar PRODUCT
Turbo Force ORG
Madison PERSON
WI ORG
Robin Vos R-Rochester PERSON
Senate ORG
Scott Fitzgerald PERSON
Evers PERSON
Evers PERSON
Wisconsin GPE
CDC ORG
1000 CARDINAL
only around 20% PERCENT

60% PERCENT
fewer than 10 CARDINAL
10% PERCENT
Evers PERSON
Memorial Day DATE
one CARDINAL
Wisconsin GPE
Northwoods GPE
Dane PERSON
Milwaukee GPE
CDC ORG
only days DATE
first ORDINAL
one day DATE
SNP ORG
Scottish Labour ORG
today DATE
The Sunday Post ORG
2017/18 CARDINAL
Scotland GPE
Scottish NORP
over 60s DATE
SNP ORG
Humza Yousaf PERSON
SNP ORG
SNP ORG
60 CARDINAL
First ORDINAL
Nicola Sturgeon PERSON
SNP ORG
Scottish Labour's Transport ORG
Neil Bibby MSP PERSON
Nicola Sturgeon ORG
Nicola Sturgeon ORG
Scotland GPE
First ORDINAL
The SNP Government ORG
10 MONEY
First ORDINAL
SNP Government ORG
The Independent ORG
six CARDINAL
10 CARDINAL
Britain GPE
Chris Grayling PERSON
thousands CARDINAL
weeks DATE
Govia Thameslink Rail GTR ORG
last week DATE
East Coast LOC

Virgin Trains PERSON
Stagecoach PERSON
BMG Research ORG
last week DATE
64 per cent MONEY
just 19 per cent MONEY
56 per cent MONEY
only 15 per cent MONEY
Inside Politics ORG
Brexit PERSON
Please ORG
Please ORG
Inside Politics ORG
CARDINAL
The Independent ORG
Inside Politics ORG
Brexit PERSON
Inside Politics ORG
CARDINAL
The Independent ORG
80 per cent MONEY
Some 64 CARDINAL
only 20 CARDINAL
More than 15,000 CARDINAL
May DATE
GCSEs ORG
Cambridgeshire GPE
Hertfordshire GPE
Last week DATE
Mr Grayling ORG
recent years DATE
Andy McDonald PERSON
Network Rail ORG
Network Rail ORG
UK GPE
50 1 CARDINAL
UK GPE
UK GPE
13 April 2021 DATE
the early hours of Tuesday morning TIME
Bamburgh Lighthouse ORG
Northumberland GPE
UK GPE
12 April DATE
2021 DATE
Rebecca Richardson PERSON
Genevieve Florence PERSON
Clissold Leisure Centre ORG
London GPE
UK GPE

11 April 2021 DATE
Fox ORG
Hill PERSON
Denmark Hill PERSON
London GPE
EPA ORG
UK GPE
10 April 2021 DATE
The Death Gun Salute ORG
the Honourable Artillery Company ORG
Britain GPE
Prince Philip GPE
the The Tower of London AFP FAC
Getty UK ORG
9 April 2021 DATE
Buckingham Palace FAC
London GPE
Britain GPE
Prince Philip GPE
Elizabeth II's PERSON
Prince Philip PERSON
more than a month DATE
April 9 DATE
Buckingham Palace FAC
99 DATE
Getty Images ORG
UK GPE
8 April 2021 DATE
James King PERSON
4:05 TIME
REUTERS ORG
UK GPE
7 April 2021 DATE
the Independent Workers' Union of Great Britain ORG
Deliveroo GPE
London GPE
UK GPE
6 April 2021 DATE
Seaham Lighthouse ORG
Durham PA GPE
UK GPE
5 April 2021 DATE
Lusamba Katalay ORG
Belly Mujinga GPE
Victoria GPE
London GPE
first ORDINAL
Belly Mujinga GPE
London GPE
Victoria GPE

UK GPE
4 April 2021 DATE
Easter Sunday at EVENT
Hengistbury Head FAC
UK GPE
3 April 2021 DATE
Bill PERSON
London GPE
EPA ORG
UK GPE
2 April 2021 DATE
Bamburgh Croquet ORG
COVID-19 ORG
Northumberland GPE
Britain GPE
UK GPE
1 April 2021 DATE
St Nicholas' Park PERSON
Warwick GPE
UK GPE
this week DATE
Easter ORG
PA UK ORG
31 March 2021 DATE
the National Covid Memorial Wall ORG
London GPE
EPA ORG
UK GPE
30 March DATE
Jamie Klingler PERSON
Reclaim These Streets ORG
Clapham Common FAC
London GPE
the Chief Inspector of Constabulary ORG
Thomas Winsor PERSON
Metropolitan Police ORG
PA UK ORG
29 March 2021 DATE
the Mersey Bowman Lawn Tennis Club ORG
Liverpool GPE
England GPE
March 29, 2021 DATE
England GPE
third ORDINAL
up to six CARDINAL
England GPE
Monday DATE
second ORDINAL
AFP ORG
Getty Images ORG

UK GPE
28 March 2021 DATE
Tenby Lifeboat ORG
Tenby GPE
Wales Reuters ORG
UK GPE
27 March 2021 DATE
Manchester Reuters ORG
UK GPE
26 March 2021 DATE
Andy Warhol PERSON
Elizabeth II PERSON
Reigning Queens ORG
1985 DATE
Elizabeth II PERSON
1985 DATE
Christie's ORG
London GPE
UK GPE
25 March 2021 DATE
Charlton Lido PERSON
London GPE
PA UK ORG
24 March 2021 DATE
Scottish Green Party ORG
Patrick Harvie PERSON
Lorna Slater PERSON
Edinburgh GPE
PA GPE
UK GPE
23 March 2021 DATE
Westbourne House ORG
Sheffield GPE
the National Reflection day EVENT
first ORDINAL
first ORDINAL
UK GPE
Tom Maddick PERSON
UK GPE
22 March 2021 DATE
Couple CARDINAL
Sean Henry PERSON
Newbiggin GPE
the-Sea LOC
Northumberland GPE
UK GPE
21 March 2021 DATE
Jersey GPE
meteor ORG
UK GPE

20 March 2021 DATE
London GPE
UK GPE
19 March 2021 DATE
St. James's Park GPE
London GPE
Britain GPE
UK GPE
18 March 2021 DATE
Highland PERSON
the National Trust's ORG
Cambridgeshire GPE
UK GPE
17 March 2021 DATE
early in the morning TIME
PA GPE
UK GPE
16 March 2021 DATE
60-metre QUANTITY
12-year-old DATE
Ansha PERSON
Ethiopia GPE
WaterAid ORG
Whitby Beach GPE
Yorkshire GPE
hours TIME
each day DATE
UK GPE
15 March 2021 DATE
2021 DATE
Teresa Solar ORG
Exchange Flags ORG
one CARDINAL
11th ORDINAL
Liverpool Biennial ORG
March 20, 2021 DATE
nine CARDINAL
PA Photo ORG
Monday March 15, 2021 DATE
PA Photo ORG
second ORDINAL
UK GPE
14 March 2021 DATE
Sarah Everard PERSON
Clapham Common FAC
Sarah Everard's PERSON
Parliament Square FAC
London AFP PERSON
Getty UK ORG
13 March 2021 DATE

Sarah Everard PERSON
Clapham Common FAC
London GPE
Getty Images ORG
UK GPE
12 March DATE
2021 DATE
Angela Rayner PERSON
the Labour Party ORG
Jonathan Ashworth PERSON
Shadow Health WORK_OF_ART
London GPE
1 percent PERCENT
Thursday DATE
May PA DATE
UK GPE
11 March 2021 DATE
Newhaven GPE
England AFP/ PERSON
UK GPE
10 March 2021 DATE
Great Chart Golf FAC
Leisure in Ashford ORG
UK GPE
9 March 2021 DATE
Kielder Water FAC
Northumberland GPE
the Border Mires ORG
50 year old DATE
the Border Mires in WORK_OF_ART
Northumberland GPE
UK GPE
8 March 2021 DATE
the Jewellery Quarter Academy ORG
Birmingham GPE
the West Midlands LOC
England GPE
first ORDINAL
two months DATE
first ORDINAL
UK GPE
7 March 2021 DATE
NHS ORG
NHS Workers ORG
Unite ORG
Downing Street FAC
London GPE
1% PERCENT
UK GPE
6 March 2021 DATE

Scottish NORP
Ibrox Stadium FAC
UK GPE
March 2021 DATE
Sutton Manor GPE
St Helens PERSON
a clear DATE
Spring day DATE
PA GPE
UK GPE
4 March 2021 DATE
Health ORG
Matt Hancock PERSON
the Centre for Virus Research ORG
Glasgow PA GPE
UK GPE
March 2021 DATE
Exchequer ORG
Rishi Sunak ORG
11 Downing Street ORG
London GPE
the House of Commons ORG
Budget PA UK ORG
2 March 2021 DATE
the Carbis Bay Estate PRODUCT
G7 CARDINAL
Getty UK ORG
1 March 2021 DATE
A J Thompson & Sons ORG
the Romney Marsh ORG
Kent PERSON
the first day DATE
UK GPE
28 CARDINAL
February 2021 DATE
Scarborough GPE
North Yorkshire GPE
UK GPE
27 February 2021 DATE
Tom Moore PERSON
the Armed Forces at Bedford Crematorium ORG
Bedford GPE
EPA ORG
UK GPE
26 February 2021 DATE
New Brighton GPE
Wirral PA UK ORG
25 February 2021 DATE
Shadow PRODUCT
Anneliese NORP

the Institute of Global Prosperity ORG
Labour Party ORG
London GPE
UK GPE
24 February 2021 DATE
Britain GPE
Health ORG
Matt Hancock PERSON
Downing Street FAC
London GPE
Britain GPE
UK GPE
23 CARDINAL
February 2021 DATE
Two CARDINAL
Boscombe GPE
Dorset PA RMT GPE
Mr Grayling ORG
Tory NORP
Network Rail ORG
GTR ORG
Network Rail ORG
Mark Carne PERSON
Queen PERSON
The Department of Transport ORG
BMG ORG
1,490 CARDINAL
Great Britain GPE
5th ORDINAL
8th June DATE
BMG ORG
the British Polling Council ORG
Australia GPE
November DATE
New South Wales GPE
thousands CARDINAL
at least 22 CARDINAL
Australia GPE
Earlier this week DATE
Bernie Sanders PERSON
Australia GPE
this bushfire season DATE
7NEWS MONEY
this bushfire season DATE
Strike Force Indarra PERSON
Dipole PRODUCT
the past decade DATE
Miranda Devine PERSON
cooler months DATE
summer DATE

I&I Editorial ORG
Joe Biden's PERSON
$2 trillion MONEY
this week DATE
one CARDINAL
Biden PERSON
Biden PERSON
his first day DATE
100% PERCENT
zero CARDINAL
House ORG
Democrats NORP
100% PERCENT
2035 DATE
today DATE
0.5% PERCENT
less than 2% PERCENT
2019 DATE
$5 billion MONEY
EV ORG
Biden PERSON
Democrats NORP
Biden PERSON
EV ORG
$2 trillion MONEY
Biden PERSON
EV ORG
Chuck Schumer PERSON
$454 billion MONEY
zero CARDINAL
one CARDINAL
zero CARDINAL
the University of Michigan's Transportation Research Institute ORG
55 miles QUANTITY
the Union of Concerned Scientists ORG
68% PERCENT
zero CARDINAL
Manhattan Institute ORG
2050 DATE
less than 500 million tons MONEY
Biden PERSON
2035 DATE
Today DATE
20% PERCENT
The Energy Information Administration ORG
32% PERCENT
2035 DATE
38% PERCENT
2050 DATE
100% PERCENT

fifth ORDINAL
today DATE
Another 43% PERCENT
the Natural Resources Defense Council ORG
Virginia GPE
490 square miles QUANTITY
Los Angeles GPE
A single CARDINAL
50 DATE
23 square miles QUANTITY
Real Clear Investigations WORK_OF_ART
Biden PERSON
U.S. GPE
zero CARDINAL
2050 DATE
the Heritage Foundation ORG
just more than half CARDINAL
Biden PERSON
Alexandria Ocasio-Cortez PERSON
Biden PERSON
Kamala Harris PERSON
Harris PERSON
The Zero Emission Vehicles Act WORK_OF_ART
2040 DATE
the I&I Editorial Board Will You Help Us In The Fight For Free Speech
ORG
the day DATE
Donations Page WORK_OF_ART
I&I ORG
Loading PERSON
Last year DATE
0.04 Celsius PERSON
2016 DATE
the hottest year DATE
The National Oceanic and Atmospheric Administration ORG
NOAA ORG
this week DATE
last year DATE
137-year DATE
2016 DATE
the third consecutive year DATE
2016 DATE
2015 DATE
0.04 degrees Celsius QUANTITY
previous years DATE
last year DATE
2015 DATE
2016 DATE
years DATE
El Niño ORG

Pacific LOC
The Cato Institute's ORG
Patrick Michaels PERSON
1998 DATE
Fahrenheit WORK_OF_ART
the following years DATE
2015 DATE
the end of 2016 DATE
2014 DATE
NOAA ORG
El Niño's ORG
Judith Curry PERSON
Congress ORG
2014 DATE
U.N. Intergovernmental Panel on Climate Change ORG
0.2 degrees Celsius QUANTITY
the early 21st century DATE
the first 15 years DATE
0.05 degrees Celsius QUANTITY
more than 40% PERCENT
1900 DATE
between 1910 and 1945 DATE
only 10% PERCENT
Maiharaji Altine PERSON
Gusau NORP
Mohammed Bello Matawalle PERSON
Zamfara State PERSON
American NORP
Zamfara PERSON
Gusau Government ORG
thousands CARDINAL
All Progressives Congress ORG
Maradun PERSON
PDP ORG
Sunday DATE
Matawalle GPE
Zamfara PERSON
one CARDINAL
Nigeria GPE
the American Government ORG
American NORP
just last week DATE
Zamfara State PERSON
Nigeria GPE
Buhari GPE
Tunde Bakare FAC
65th ORDINAL
within few months DATE
Matawalle PERSON
APC ORG

the Peoples Democratic Party ORG
APC ORG
Aminu Gamagiwa ORG
APC ORG
Maradun PERSON
PDP ORG
Matawalle PERSON
Gamagiwa PERSON
24,000 CARDINAL
APC ORG
Maradun PERSON
PDP ORG
Sunday DATE
U.S. GPE
Kamala Harris PERSON
Harris PERSON
Oct. 20, 1964 DATE
Donald Jasper Harris PERSON
Jamaica GPE
Gopalan Shyamala PERSON
India GPE
Jamaica GPE
India GPE
Kamala Harris PERSON
Jamaica GPE
Section 3Cb LAW
India GPE
II CARDINAL
Sec ORG
5 CARDINAL
Tyla PERSON
Douglas PERSON
Michael PERSON
Kamala Harris PERSON
United States GPE
1,200 MONEY
Trump PERSON
millions CARDINAL
Democrats NORP
Social Security ORG
Democrats NORP
Nancy Pelosi's ORG
Joe Barron PERSON
1,200 MONEY
4 days DATE
max PERSON
one cent MONEY
Donald Trump PERSON
Democrats NORP
November DATE

Greenland GPE
annual DATE
two CARDINAL
Johan Petersen PERSON
Greenland GPE
six metres QUANTITY
annual DATE
Greenland GPE
Greenland GPE
six metres QUANTITY
Greenland GPE
20,000 years ago DATE
GPS ORG
Greenland GPE
12mm QUANTITY
19 cubic kilometres QUANTITY
each year DATE
about 8% PERCENT
Greenland GPE
Greenland GPE
40m years ago DATE
Iceland GPE
Greenland GPE
millennia GPE
Greenland GPE
the past and today DATE
the last 20 years DATE
Greenland GPE
decades DATE
Prof Jonathan Bamber PERSON
the University of Bristol ORG
UK GPE
one CARDINAL
Science Advances ORG
One CARDINAL
Dr Christopher Harig PERSON
the University of Arizona ORG
the end of the century DATE
years ago DATE
Pippa Whitehouse PERSON
the University of Durham ORG
GPS ORG
Greenland GPE
Melt PERSON
Greenland GPE
10 June, 2014 DATE
15 June, 2016 DATE
Every spring DATE
early summer DATE
one CARDINAL

2016 DATE
Antarctica LOC
April DATE
Greenland GPE
August DATE
GETTY ORG
Forty-seven DATE
SUBSCRIBE Invalid PRODUCT
HM Inspectorate of Prisons ORG
annual DATE
2001/2002 DATE
2015/16 CARDINAL
47per cent MONEY
2001 DATE
37per cent MONEY
Muslim NORP
22per cent MONEY
19per cent MONEY
Romany PERSON
YOI ORG
41per cent MONEY
53per cent MONEY
2010-11 DATE
2015-16 DATE
HM Inspectorate of Prisons ORG
April DATE
906 CARDINAL
England GPE
Wales GPE
Peter Clarke PERSON
HMIP ORG
the Youth Justice Board ORG
18-year-olds' DATE
nearly half CARDINAL
46per cent MONEY
the past 12 months DATE
28per cent MONEY
2014-15 DATE
Fri, June 17, 2016 DATE
Getty Images 1 PERSON
30 CARDINAL
one CARDINAL
three CARDINAL
El Salvador GPE
53 per cent MONEY
between 2010-11 DATE
2015-16 DATE
Nearly half CARDINAL
Democrats NORP
Democrat NORP

AOC ORG
Social Security ORG
the age of 75 DATE
second ORDINAL
Republicans NORP
Republicans NORP
Joe Barron PERSON
next month DATE
AOC ORG
first ORDINAL
years DATE
Republicans NORP
AOC ORG
AOC ORG
first ORDINAL
n't GPE
our year 2020 DATE
China GPE
Severe Fever ORG
three years ago DATE
16 CARDINAL
summer DATE
China GPE
two CARDINAL
hours TIME
Benjamin H. Strauss PERSON
one CARDINAL
two CARDINAL
Monday DATE
second ORDINAL
28 centuries DATE
the past century DATE
the next few decades DATE
as much as three or CARDINAL
2100 DATE
the 22nd century DATE
36 percent PERCENT
1.36 CARDINAL
1.51 CARDINAL
Table 5 LAW
Influenza NORP
Department of Defense ORG
2017—2018 influenza season DATE
Flu Vaccine Interference Paraphrasing PERSON
the Department of Veterans Affairs' ORG
two weeks ago DATE
one CARDINAL
PubMed ORG
Informed Consent For ORG
Flu Vaccine Now PERSON

Table 5 LOC
Table 5 LAW
Coronavirus GPE
1.36 DATE
36% PERCENT
0.57 CARDINAL
Three CARDINAL
Three CARDINAL
8 CARDINAL
36% PERCENT
1 MONEY
this year DATE
Americans NORP
VA ORG
Vaccine PERSON
January 2020 DATE
the Armed Forces Health Surveillance Branch Air Force Satellite ORG
Wright-Patterson AFB ORG
OH GPE
the Air Force Research Laboratory Institutional Review Board ORG
VA ORG
1151 DATE
the US Court of Claims — Vaccine Court ORG
US GPE
DOD ORG
VA ORG
Veterans PERSON
VA GPE
Vietnam GPE
Vietnam GPE
VA ORG
the Department of Defense Global Respiratory Pathogen Surveillance
Program DoDGRS ORG
the Global Emerging Infections Surveillance ORG
2017-2018 DATE
2017—2018 CARDINAL
1 October 2017 and ended 29 September 2018 DATE
Chlamydia ORG
Mycoplasma GPE
non-influenza NORP
the season DATE
the season DATE
US GPE
YouTube ORG
60 CARDINAL
Australia GPE
Wuhan GPE
4 CARDINAL
Rainsy PERSON
Cambodia GPE

Sam Rainsy C PERSON
L GPE
seven CARDINAL
CNRP ORG
Cambodia GPE
Sam Rainsy C PERSON
L GPE
seven CARDINAL
CNRP ORG
Facebook Government ORG
the Phnom Penh Municipal Court's ORG
Sam Rainsy PERSON
eight CARDINAL
Ministry of Justice ORG
Chin Malin PERSON
yesterday DATE
Duch Sok Sarin PERSON
Monday DATE
Rainsy PERSON
25 years DATE
Mu Sochua PERSON
Eng Chhai Eang PERSON
22 years DATE
six CARDINAL
20 years DATE
Rainsy PERSON
Tioulong Saumura PERSON
Ou Chanrith PERSON
Ho Vann PERSON
Men Thavarin GPE
Nuth Romduol ORG
Kingdom GPE
November 9, 2019 DATE
Rainsy PERSON
Cambodian NORP
Hun Sen PERSON
Cambodian NORP
Rainsy PERSON
Hun Sen PERSON
Cambodia GPE
CNRP ORG
Malin PERSON
Malin PERSON
CNRP ORG
Cambodia GPE
130 CARDINAL
the past year DATE
nine CARDINAL
CNRP ORG
The US Embassy ORG

Phnom Penh GPE
yesterday DATE
nine CARDINAL
Cambodian NORP
American NORP
more than 130 CARDINAL
Cambodia GPE
2022 DATE
Cambodian People's Party ORG
Sok Eysan PERSON
yesterday DATE
CNRP ORG
2022 DATE
2023 CARDINAL
Commune and National Elections ORG
Bacon ORG
The World Health Organisation ORG
The World Health Organisation ORG
WHO ORG
more than 150,000 CARDINAL
UK GPE
Britain GPE
second ORDINAL
half CARDINAL
The Department of Health's ORG
WHO ORG
Monday DATE
ten CARDINAL
UK GPE
five CARDINAL
five CARDINAL
Barry Carpenter PERSON
the North American Meat Institute ORG
pastrami GPE
one CARDINAL
one CARDINAL
30 CARDINAL
20 CARDINAL
day DATE
English LANGUAGE
almost 20,000 CARDINAL
UK GPE
each year DATE
The World Health Organisation ORG
asbesto Any PERSON
last year DATE
WHO ORG
2.8billion MONEY
440,000 CARDINAL
England GPE

The North American Meat Institute NAMI ORG
dozens CARDINAL
940 CARDINAL
International Agency for Research ORG
Barry Carpenter PERSON
NAMI ORG
IARC NORP
Richard Knox PERSON
Institute of Cancer Research ORG
2011 DATE
no more than 70 CARDINAL
Evans PERSON
Dueker PERSON
St. Louis GPE
first ORDINAL
four years DATE
Cardinals ORG
Scott Faughn PERSON
Missouri GPE
Eric Schmitt PERSON
This Week DATE
Missouri Politics ORG
Missouri Republican Party ORG
Jean Evans PERSON
Jane Dueker PERSON
St. Louis GPE
Missouri GPE
Missouri GPE
Evans PERSON
Last week DATE
Republican NORP
Mike Parson PERSON
The same week DATE
St. Louis GPE
Evans ORG
Parson ORG
Evans PERSON
Dueker PERSON
one CARDINAL
Dueker PERSON
Ten percent PERCENT
90 percent PERCENT
Donald Trump PERSON
HD 99 DATE
Washington GPE
U.S. GPE
David Cicilline PERSON
the Paycheck Fairness Act LAW
House ORG
today DATE

House ORG
Republicans NORP
Nearly a century DATE
only 77 cents MONEY
Cicilline. PERSON
January 2009 DATE
Democratic NORP
111th ORDINAL
Congress ORG
first ORDINAL
Obama PERSON
House ORG
Republicans NORP
January 2011 DATE
the Paycheck Fairness Act LAW
the Equal Pay Act LAW
House ORG
Republican NORP
Rhode Island GPE
Cicilline PRODUCT
This year DATE
Tax Day DATE
Cicilline PRODUCT
Cicilline PERSON
Boehner PERSON
last week DATE
House ORG
Buffett Rule WORK_OF_ART
today DATE
multimillion dollar MONEY
Rhode Island GPE
Microsoft ORG
Bill Gates PERSON
COVID-19 PERSON
700,000 CARDINAL
10,000 CARDINAL
700,000 CARDINAL
's GPE
Europe LOC
kimchi GPE
South Korea GPE
Asian NORP
COVID-19 ORG
South Korea GPE
France GPE
Pulmonary Medicine PERSON
Montpellier University ORG
Jean Bousquet PERSON
ACE2 PRODUCT
COVID-19 ORG

ACE2 PRODUCT
Sun ORG
South Korea GPE
296 CARDINAL
13,771 CARDINAL
COVID-19 ORG
2.14% PERCENT
Read More Advertisement   Meanwhile WORK_OF_ART
Germany GPE
4.51% PERCENT
9,163 CARDINAL
202,845 CARDINAL
Italy GPE
14.33% PERCENT
Spain GPE
9.24% PERCENT
the United Kingdom GPE
15.36% PERCENT
Greece GPE
Bulgaria GPE
Turkey GPE
COVID-19 PERSON
Dr Bousquet PERSON
the WHO Global Alliance ORG
the University of Leipzig ORG
Germany GPE
April DATE
the New York Post ORG
kimchi GPE
New York-Presbyterian GPE
Weill Cornell Medicine Division of Gastroenterology and Hepatology ORG
Robert Brown PERSON
Brown PERSON
COVID-19 PERSON
January DATE
South Korea's GPE
Health Ministry ORG
kimchi GPE
Korean NORP
one hundred thousand CARDINAL
the Royal Family ORG
Palace ORG
Government ORG
the Royal Family ORG
£369 million MONEY
Buckingham Palace FAC
Palace ORG
the Second World War EVENT
the Royal Family ORG
First ORDINAL

Buckingham Palace FAC
Queen PERSON
Monarch LOC
the Crown Estate ORG
HM Treasury ORG
The Monarch of the day EVENT
Palace ORG
hundreds CARDINAL
every year DATE
Secondly ORDINAL
the Crown Estate ORG
six CARDINAL
the last ten years DATE
£2.4 billion MONEY
Treasury ORG
NHS ORG
the last year alone DATE
£304.1 million MONEY
Exchequer ORG
Crown Estate PRODUCT
15% PERCENT
the Crown Estate ORG
over 85% PERCENT
Queen PERSON
Palace ORG
the Crown Estate ORG
less than two years DATE
Queen PERSON
every year DATE
British NORP
Second ORDINAL
Swedish NORP
Swedish NORP
Swedes NORP
Earlier this month DATE
U.K. GPE
Swedes NORP
Swedish NORP
Swedish Microchip Procedures Pictures Flood Instagram Sandra
Wuerthner ORG
Wuerthner ORG
Wuerthner ORG
Swedish NORP
Microchipping Is Voluntary Corporations WORK_OF_ART
Jim Satney PrepForThat PERSON
Amazon ORG
CDC ORG
COVID-19 ORG
Amazon ORG
Amazon ORG

Amazon ORG
As much as 40% PERCENT
Amazon ORG
Nature Communications ORG
decades DATE
Rainforests NORP
Amazon ORG
about 40% PERCENT
the Stockholm Resilience Centre ORG
Last year DATE
Brazil GPE
Jair Bolsonaro PERSON
Amazon ORG
year DATE
Amazon ORG
a decade DATE
60% PERCENT
last year DATE
Arie Staal PERSON
Forests NORP
Staal ORG
Guardian ORG
Amazon ORG
the end of this century DATE
Amazon ORG
the end of this century DATE
early 21st century DATE
climate(2003-2014 DATE
late 21st century DATE
2071-2100)Climate DATE
Guardian ORG
al, Nature Communications ORG
Obbe Tuinenburg PERSON
earth LOC
Ingo Fetzer PERSON
the Stockholm Resilience Centre ORG
many decades DATE
1 CARDINAL
Saint Francis PERSON
Assisi GPE
earth LOC
2:7 CARDINAL
3 CARDINAL
More than fifty years ago DATE
Saint John XXIII PERSON
Encyclical ORG
Terris GPE
Catholic NORP
Church ORG
Encyclical ORG

4. CARDINAL
1971 DATE
eight years DATE
Pacem PERSON
Terris GPE
Blessed Pope PERSON
Paul VI PERSON
the Food and Agriculture Organization of the United Nations ORG
Saint John Paul II PERSON
Encyclical ORG
today DATE
6 CARDINAL
Benedict XVI PERSON
Benedict PERSON
Benedict PERSON
7 CARDINAL
Church ORG
the Catholic Church ORG
Churches PRODUCT
Christian NORP
Ecumenical Patriarch Bartholomew ORG
8 CARDINAL
Patriarch Bartholomew PERSON
earth LOC
Bartholomew PERSON
Christians NORP
Saint Francis PERSON
Assisi GPE
Encyclical ORG
Bishop ORG
Rome GPE
Saint Francis PERSON
non-Christians NORP
11 CARDINAL
Francis PERSON
Saint Bonaventure GPE
Saint Francis PERSON
12 CARDINAL
Saint Francis PERSON
Wis 13:5 PRODUCT
Rom 1:20 DATE
Francis PERSON
13 CARDINAL
Southern Africa LOC
22 CARDINAL
15 CARDINAL
Encyclical Letter ORG
Church ORG
today DATE
Judaeo-Christian NORP

Christian NORP
16. CARDINAL
Encyclical ORG
CHAPTER ORG
17 CARDINAL
18 CARDINAL
today DATE
today DATE
20 CARDINAL
daily DATE
millions CARDINAL
one CARDINAL
21 CARDINAL
Each year DATE
hundreds of millions of tons MONEY
22 CARDINAL
one CARDINAL
one CARDINAL
23 CARDINAL
recent decades DATE
earth LOC
recent decades DATE
earth LOC
24 CARDINAL
this century DATE
a quarter CARDINAL
25 CARDINAL
one CARDINAL
our day DATE
coming decades DATE
26 CARDINAL
the next few years DATE
27 CARDINAL
28 CARDINAL
Africa LOC
29 CARDINAL
One CARDINAL
30 CARDINAL
31 CARDINAL
a few decades DATE
billions CARDINAL
32 CARDINAL
earth LOC
years DATE
33 CARDINAL
thousands CARDINAL
thousands CARDINAL
34 CARDINAL
35 CARDINAL
36 CARDINAL

Jim Francesconi PERSON
the State Board of Higher Education ORG
the University of Oregon ORG
Portland State University ORG
Oregon State University ORG
three CARDINAL
Stanford ORG
Yale ORG
Harvard ORG
Oregon GPE
Oregon GPE
Oregon GPE
more than 20 years DATE
Portland ORG
Oregon GPE
ONAMI ORG
Oregon Nanoscience ORG
Microtechnologies Institute ORG
the Portland Life Sciences ORG
the Light Metals Institute ORG
Oregon GPE
Washington GPE
21 CARDINAL
Oregon GPE
almost 10 percent PERCENT
Oregonians NORP
Latino PERSON
Oregon GPE
America GPE
Oregon GPE
Nearly half CARDINAL
Oregon GPE
Half CARDINAL
Oregon GPE
Pell Grants PERSON
more than $25,000 MONEY
Oregonians NORP
first ORDINAL
first ORDINAL
Oregon GPE
Oregonians NORP
UO GPE
PSU ORG
OSU ORG
Southern, Western ORG
Eastern ORG
the Oregon Institute of Technology ORG
Oregonians NORP
Oregonians NORP
Legislature ORG

Oregon GPE
Jim Francesconi PERSON
the State Board of Higher Education ORG
Northeast Portland GPE
dozens CARDINAL
the United States GPE
a quarter century DATE
Gernot Wagner PERSON
Harvard University ORG
the Environmental Defense Fund ORG
Greenland GPE
the Dust Bowl EVENT
Midwest LOC
Hsiang PERSON
Science ORG
the United States GPE
America GPE
70 years ago DATE
the United States GPE
U.S. County GPE
2080–2099 CARDINAL
Kopp PERSON
Hsiang PERSON
al. / Science ORG
half CARDINAL
the Gulf of Mexico LOC
20-percent PERCENT
Harvests ORG
summer DATE
Jonathan Ashworth PERSON
Labour's Shadow Health ORG
today DATE
3-4 months DATE
Labour ORG
Britain GPE
3-4 months DATE
Labour ORG
an additional £25 million MONEY
Labour ORG
National Child Health Fund ORG
Britain GPE
Today DATE
Labour ORG
Jonathan Ashworth PERSON
Health Visitor Programme ORG
2015 DATE
Between 2016/17 CARDINAL
2017/18 DATE
£55.2 million MONEY
0 CARDINAL

this Tory Government ORG
8,244 CARDINAL
NHS ORG
December 2017 DATE
August 2013 DATE
10,309 CARDINAL
October 2015 DATE
more than 2,065 CARDINAL
20% PERCENT
just over two years DATE
12% PERCENT
a New Birth Visit ORG
17% PERCENT
the South West LOC
England GPE
17% PERCENT
6-8 week DATE
London GPE
1/3 CARDINAL
25% PERCENT
one year DATE
12 CARDINAL
London GPE
44% PERCENT
90% PERCENT
North East LOC
two and a half year DATE
just 64% PERCENT
London GPE
73% PERCENT
the East of England LOC
Jonathan Ashworth PERSON
today DATE
Infant Feeding PERSON
a Labour Government ORG
the Coalition Government ORG
UNICEF ORG
Scotland GPE
Northern Ireland GPE
70% PERCENT
the year 2030 DATE
Jonathan Ashworth PERSON
Labour's Shadow Health ORG
London GPE
Michael Marmot PERSON
UK GPE
70th ORDINAL
NHS ORG
David Cameron PERSON
Theresa May PERSON

more than 20 percent PERCENT
just over two years DATE
England GPE
UK GPE
Health Visitors ORG
3-4 months DATE
£25 million MONEY
National Child Health Fund ORG
20 per cent MONEY
100 per cent MONEY
California GPE
California GPE
Congress ORG
late Tuesday DATE
2013 DATE
Times ORG
Oct. 25 DATE
Parsons Brinckerhoff PERSON
October 2013 DATE
first ORDINAL
31% PERCENT
2014 DATE
four months later DATE
A dozen CARDINAL
Congress ORG
four CARDINAL
California GPE
Brinckerhoff PERSON
Dan Richard PERSON
Jeff Morales PERSON
Richard PERSON
Assembly ORG
Toni Atkins PERSON
Jeff Denham PERSON
R-Turlock ORG
House ORG
Parsons Brinckerhoff PERSON
Tuesday DATE
the California High Speed Rail Authority ORG
11 CARDINAL
Republicans NORP
first ORDINAL
Burbank ORG
Merced ORG
$27.3 billion to $35.7 billion MONEY
$31 billion MONEY
about $40 billion MONEY
Brinckerhoff PERSON
Advertisement Interested PERSON
California GPE

Essential GPE
California GPE
about 5% PERCENT
$54.4 billion to $56.9 billion MONEY
Brinckerhoff PERSON
$68 billion MONEY
more than $71 billion MONEY
Brinckerhoff PERSON
Times ORG
PowerPoint ORG
22 CARDINAL
Richard PERSON
Times ORG
the San Gabriel FAC
Tehachapi GPE
Times ORG
about 36 miles QUANTITY
2022 DATE
Times ORG
Brinckerhoff PERSON
the California Public Records Act ORG
Dennis Trujillo PERSON
Times ORG
House ORG
Republicans NORP
Gregory Kelly ORG
Parsons Brinckerhoff PERSON
Parsons Brinckerhoff PERSON
the Rail Authority ORG
the California Assembly ORG
Jim Patterson PERSON
Fresno GPE
Atkins PERSON
Patterson PERSON
Patterson PERSON
Richard PERSON
Fresno GPE
Tuesday DATE
$68-billion MONEY
Richard PERSON
Denham PERSON
Denham PERSON
the Republican Party ORG
House ORG
Senate ORG
$16.3-billion MONEY
2022 DATE
@rvartabedian GPE
Measure WORK_OF_ART
San Francisco GPE

Ed Lee PERSON
Advertisement Pit PERSON
Thor PERSON
Texas GPE
California GPE
Alibaba BABA PERSON
Microsoft MSFT ORG
Stanford University ORG
first ORDINAL
Alibaba PERSON
Monday DATE
Stanford ORG
Alibaba PERSON
first ORDINAL
Google ORG
China GPE
Luo Si PERSON
Chinese NORP
AI ORG
Si PERSON
Alibaba GPE
Singles Day DATE
Pranav Rajpurkar PERSON
one CARDINAL
Stanford ORG
Alibaba ORG
2018 DATE
three CARDINAL
Stanford ORG
Wikipedia ORG
AI ORG
British NORP
TARDIS ORG
Alibaba PERSON
82.44 CARDINAL
January 11 DATE
82.304 CARDINAL
A day later DATE
Microsoft ORG
AI ORG
82.650 CARDINAL
AI ORG
Andrew Pickup PERSON
Microsoft ORG
AI ORG
Facebook FB ORG
TCEHY ORG
Samsung ORG
AI ORG
Stanford ORG

Jack Ma PERSON
Russian NORP
Vladimir Putin PERSON
September DATE
China GPE
Beijing GPE
2020 DATE
July DATE
nearly $150 billion MONEY
the next couple years DATE
NHS ORG
England GPE
more than £1 billion MONEY
NHS Digital ORG
more than £422 million MONEY
the last 10 years DATE
Almost one CARDINAL
20 CARDINAL
2 CARDINAL
around 90% PERCENT
Robin Hewings PERSON
UK GPE
the last 20 years DATE
26,000 CARDINAL
1bn ORDINAL
NHS ORG
Hewings PERSON
2 CARDINAL
years earlier' DATE
Postcode PERSON
Presentational NORP
over three million CARDINAL
England GPE
the last two decades DATE
nearly 100,000 CARDINAL
92% PERCENT
2 CARDINAL
2 CARDINAL
1 CARDINAL
Almost seven CARDINAL
10 CARDINAL
66.8% PERCENT
almost six CARDINAL
10 CARDINAL
57.8% PERCENT
England GPE
2 CARDINAL
more than 80 percent PERCENT
obese NORP
2 CARDINAL

NHS England LOC
Nearly £477 million MONEY
2017-18 DATE
the same year DATE
around £350 million MONEY
£181 million MONEY
Jonathan Valabhji PERSON
NHS England FAC
NHS ORG
2 CARDINAL
first ORDINAL
The NHS Diabetes Prevention Programme ORG
million CARDINAL
2 CARDINAL
three-year-old DATE
last night TIME
three-year-old DATE
July 31 DATE
58 DATE
70 DATE
two CARDINAL
each year DATE
18 CARDINAL
Last night TIME
two days DATE
these days DATE
100 per cent MONEY
Grandfather PERSON
70 DATE
Britain GPE
Sue Tollefsen PERSON
Freya PERSON
2008 DATE
57 DATE
30s DATE
January DATE
One week later DATE
Shoeburyness GPE
Essex GPE
March DATE
THE BIAS OF SOCIAL ORG
The Children Act LAW
Special Guardianship ORG
Last year DATE
two-year-old DATE
Chelmsford ORG
June 17 DATE
the Daily Mail ORG
82 CARDINAL
70 CARDINAL

15 CARDINAL
yesterday DATE
Southend Council ORG
Karina Chetwynd PERSON
John Copland PERSON
Son NORP
Chetwynd PERSON
Anne Jones PERSON
ROME ORG
Reuters ORG
Catholic NORP
Italy GPE
these days DATE
St. Mark's Square GPE
Italy GPE
Venice GPE
Italy GPE
March 10, 2020 DATE
REUTERS ORG
Manuel Silvestri PERSON
Italy GPE
Catholic Italy ORG
Europe LOC
Last Saturday DATE
Italian NORP
first ORDINAL
Sunday DATE
Riccardo Lamba PERSON
San Ponziano GPE
Rome GPE
About 70 CARDINAL
2 CARDINAL
Lamba PERSON
this Sunday DATE
the day DATE
Tuesday DATE
Basilica LOC
Tuesday DATE
Vatican ORG
at least April 3 DATE
a matter of hours TIME
Lamba ORG
at least April 3 DATE
Sabrina Bucci PERSON
Esclusivevent GPE
Rome GPE
St. Peter's Basilica GPE
five CARDINAL
Catholics NORP
morning TIME

Stefano Chiericoni PERSON
one CARDINAL
Rome GPE
The National Oceanic and Atmospheric Administration ORG
NOAA ORG
U.S. GPE
NOAA ORG
NOAA ORG
winter DATE
NOAA ORG
January 2018 DATE
Northeast U.S. LOC
New England LOC
NY ORG
PA GPE
NJ ORG
DE GPE
MD GPE
Paul Homewood PERSON
NOAA ORG
NOAA ORG
3.1 degrees QUANTITY
first ORDINAL
New York GPE
NOAA ORG
winter DATE
New York GPE
2013/14 CARDINAL
Jan 2nd DATE
Arctic LOC
March DATE
NWS ORG
the end of the winter DATE
The winter of DATE
2013 CARDINAL
New York State GPE
Western NORP
North Central New York ORG
This winter DATE
two CARDINAL
winters DATE
this winter DATE
this winter DATE
normal every month DATE
January through March DATE
at least 4 CARDINAL
two CARDINAL
Western New York ORG
Buffalo GPE
Rochester GPE

the month of January DATE
NOAA ORG
30th ORDINAL
1895 DATE
New York State GPE
16.9F CARDINAL
January 1943 and January 2014 — months DATE
NOAA ORG
NOAA ORG
2014 DATE
NOAA ORG
1943 DATE
Jan 2014 DATE
2.7F CARDINAL
1943 DATE
NOAA ORG
NOAA ORG
Homewood GPE
Big Freeze PERSON
the winter of 2017/2018 DATE
NOAA ORG
January 2018 DATE
New York GPE
January 1943 DATE
three CARDINAL
Ithaca GPE
Auburn FAC
Geneva GPE
January 2018 DATE
January 1943 DATE
1.0 CARDINAL
1.7 DATE
1.3F CARDINAL
NOAA ORG
2.1F warmer last month DATE
NOAA ORG
last month DATE
1943 DATE
NOAA ORG
Central Lakes LOC
one CARDINAL
US GPE
NOAA ORG
One CARDINAL
NOAA ORG
Syracuse GPE
1929 DATE
Homewood ORG
two CARDINAL
two million CARDINAL

the Urban Heat Island ORG
NOAA ORG
NOAA ORG
U.S. GPE
Marijuana PERSON
LEGAL ORG
marijuana PERSON
marijuana PERSON
marijuana PERSON
Marijuana Use PERSON
the United States GPE
Marijuana PERSON
the National Institute of Drug Abuse ORG
18 to 25 DATE
more than half CARDINAL
marijuana PERSON
two CARDINAL
16.40 percent PERCENT
12 to 17 DATE
45.70 percent PERCENT
26 DATE
marijuana PERSON
1 CARDINAL
Laws in more than half WORK_OF_ART
US GPE
marijuana PERSON
2015 DATE
the District of Columbia GPE
marijuana PERSON
2 CARDINAL
marijuana PERSON
Trip ORG
2014 DATE
marijuana PERSON
3 CARDINAL
2015 DATE
Degenhardt PERSON
al. PERSON
marijuana PERSON
4 CARDINAL
marijuana PERSON
marijuana PERSON
One CARDINAL
mid-2014 DATE
marijuana PERSON
5 CARDINAL
marijuana PERSON
marijuana PERSON
marijuana PERSON
the United States GPE

January 2015 DATE
Lechenmeier ORG
Rehm PERSON
6 CARDINAL
the Margin of Exposure MOE ORG
two CARDINAL
1 CARDINAL
2 CARDINAL
MOE ORG
MOE ORG
MDMA ORG
THC ORG
marijuana PERSON
100 CARDINAL
marijuana PERSON
National Institute of Drug Abuse ORG
2014 DATE
Marijuana PERSON
http://www.drugabuse.gov/drugs-abuse/marijuana ORG
2 CARDINAL
R. NORP
2015 DATE
Marijuana Laws PERSON
Devil PERSON
3 CARDINAL
Tripp GPE
D. NORP
2014 DATE
one year DATE
the Olympic Winter Games Beijing 2022 EVENT
the International Olympic Committee IOC ORG
Athletes GPE
between 4 and 20 CARDINAL
next year's DATE
IOC ORG
Thomas Bach PERSON
today DATE
winter DATE
next year's DATE
IOC ORG
the Olympic Games EVENT
today DATE
Olympic House ORG
Lausanne GPE
Switzerland GPE
IOC ORG
eight CARDINAL
the Chinese Olympic Committee ORG
seven CARDINAL
the Hellenic Olympic Committee ORG

the Olympic Games EVENT
Olympic Games EVENT
the Japanese Olympic Committee for Tokyo 2020 ORG
the Comité National Olympique et Sportif Français ORG
Paris GPE
2024 DATE
the Comitato Olimpico Nazionale Italiano for Milano Cortina 2026 ORG
Beijing GPE
the United States GPE
the Australian Olympic Committee ORG
the Nigerian Olympic Committee ORG
Oceania GPE
Africa LOC
Olympic Winter Games EVENT
eight CARDINAL
Bach PERSON
One year from now DATE
Beijing GPE
first ORDINAL
the summer DATE
winter DATE
the Olympic Games EVENT
Chinese NORP
China GPE
Chinese NORP
Olympic Games EVENT
China GPE
the Olympic Winter Games Beijing 2022 EVENT
China GPE
winter DATE
Olympic Winter Games Beijing 2022 EVENT
StrongerTogether PERSON
IOC ORG
next year's DATE
Chinese NORP
Eileen Gu PERSON
Olympic Winter Games EVENT
Beijing 2022's ORG
Beijing GPE
@Olympics ORG
StrongerTogether PERSON
emoji PERSON
Beijing 2022 EVENT
Snapchat PERSON
daily DATE
the Olympic Channel and @Olympics Twitter ORG
the Olympics Instagram ORG
Worldwide Olympic Partners ORG
the coming days DATE
Beijing GPE

OMEGA ORG
the Official Timekeeper of the Olympic Games ORG
winter DATE
Beijing GPE
2022 DATE
Olympian NORP
Paralympian Online Experiences ORG
Worldwide Olympic ORG
Paralympic Partner Airbnb ORG
Omega PERSON
China GPE
One Year DATE
the Beijing 2022 Organising Committee ORG
Beijing 2022 EVENT
the Water Cube WORK_OF_ART
next year's DATE
Paralympic Winter Games EVENT
1YearToGo MONEY
February 4 DATE
next year DATE
China GPE
the end of 2020 DATE
the end of this year DATE
300 million CARDINAL
winter DATE
one million CARDINAL
the Beijing 2022 Games ORG
China GPE
winter DATE
winter DATE
the Olympic Winter Games Beijing 2022 EVENT
winter DATE
Five years ago DATE
460 FAC
China GPE
644 CARDINAL
the end of 2019 DATE
The Olympic Winter Games Beijing 2022 EVENT
100 per cent MONEY
CO 2 ORG
Beijing GPE
first ORDINAL
China GPE
the Olympic Winter Games EVENT
7 CARDINAL
12 CARDINAL
Beijing 2022 EVENT
Beijing GPE
2008 DATE
Beijing GPE

Olympic Winter Games EVENT
4 CARDINAL
February 2022 DATE
MONEY
The International Olympic Committee ORG
more than 90 CARDINAL
USD 3.4 million CARDINAL
MONEY
the IOC Media Relations Team ORG
21 CARDINAL
6000 CARDINAL
Videos GPE
Flickr LOC
Images NORP
IOC ORG
Twitter PRODUCT
YouTube ORG
Obama GPE
U.S. GPE
State ORG
John Kerry PERSON
Iranian NORP
Javad Zarif PERSON
Vienna GPE
July 2015 DATE
WASHINGTON GPE
Obama GPE
$400 million MONEY
Iran GPE
January DATE
four CARDINAL
Americans NORP
Tehran GPE
U.S. GPE
European NORP
Iran GPE
U.S. GPE
Netherlands GPE
Switzerland GPE
first ORDINAL
$1.7 billion MONEY
Obama GPE
Iran GPE
decades-old DATE
the 1979 fall DATE
Iran GPE
Shah Mohammad Reza Pahlavi PERSON
The Hague GPE
Tehran GPE
U.S. GPE

summer DATE
Barack Obama PERSON
the White House FAC
Jan. 17 DATE
$400 million MONEY
Democrats NORP
the House of Representatives ORG
Republicans NORP
US GPE
Russia GPE
Michael McFaul PERSON
Russia GPE
McFaul PERSON
two years DATE
Barack Obama's PERSON
Moscow GPE
Twitter PRODUCT
congress ORG
Russian LANGUAGE
Russian NORP
Vladimir Putin PERSON
Michael McFaul PERSON
US GPE
Russia GPE
McFaul PERSON
Russian NORP
Vladimir Vladimirovich Putin PERSON
Russian NORP
Democrats NORP
Привет ВВП. PERSON
Michael McFaul PERSON
November 7, 2018 DATE
Democrats NORP
Donald Trump PERSON
Russia GPE
McFaul PERSON
Russia GPE
McFaul ORG
Russia GPE
US GPE
Russia GPE
Twitter PRODUCT
Trump ORG
McFaul PERSON
Russia GPE
Putin PERSON
US GPE
McFaul PERSON
Putin PERSON
Trump ORG

Trump ORG
Russia GPE
Russia GPE
European NORP
Moscow GPE
McFaul PERSON
Russian NORP
last night TIME
Bill Browder PERSON
US GPE
British NORP
Putin PERSON
anti-Putin PERSON
Russia GPE
Putin PERSON
Orange County GPE
Dana Rohrabacher PERSON
Rohrabacher PERSON
US GPE
Russia GPE
Kremlin ORG
Putin PERSON
Orange County GPE
Dana Rohrabacher PERSON
Harley Rouda ORG
Rohrabacher PERSON
Congress ORG
the Magnitsky Act LAW
Bill Browder PERSON
November 7, 2018 DATE
HON ORG
STEFAN I. MYCHAJLIW ERIE COUNTY COMPTROLLER'S PERSON
AUDIT & CONTROL ORG
95 CARDINAL
FRANKLIN STREET BUFFALO ORG
NEW YORK GPE
14202 DATE
January 29 DATE
2021 DATE
Erie County Legislature ORG
Franklin Street FAC
4th Floor Buffalo ORG
New York GPE
14202 DATE
The Erie County Comptroller's Office ORG
252 CARDINAL
CARES Act LAW
252 CARDINAL
252 CARDINAL
252 CARDINAL

252 CARDINAL
August DATE
252 CARDINAL
March 25, 2020 DATE
United States GPE
Senate ORG
Coronavirus Aid, Relief and Economic Security Act WORK_OF_ART
CARES ORG
United States GPE
House of Representatives ORG
the following day DATE
March 27, 2020 DATE
The CARES Act LAW
COVID-19 ORG
Erie County GPE
$160 million MONEY
Erie County GPE
CARES Act PRODUCT
252 CARDINAL
252 CARDINAL
19 CARDINAL
COVID-19 ORG
COVID-19 ORG
COVID-19 ORG
CARES Act LAW
COVID-19 ORG
COVID-19 ORG
daily DATE
December 31, 2020 DATE
December, 2020 DATE
2020 DATE
December 31, 2021 DATE
252 CARDINAL
252 CARDINAL
COVID-19 ORG
252 CARDINAL
252 CARDINAL
CARES Act LAW
the United States Treasury Department ORG
the CARES Act LAW
January 15, 2021 DATE
Coronavirus Relief Fund Guidance ORG
Tribal Governments PRODUCT
The CARES Act LAW
1 CARDINAL
the Coronavirus Disease LOC
2019 DATE
2 CARDINAL
March 27, 2020 DATE
the CARES Act LAW

State ORG
3 CARDINAL
March 1, 2020 DATE
December 30, 2021 DATE
CARES Act LAW
December 31, 2020 DATE
Division N ORG
2021 DATE
The Department of the Treasury ORG
Overtime For the ORG
252 CARDINAL
January 1, 2021 DATE
January 8, 2021 DATE
1,226 CARDINAL
114,524 hours TIME
252 CARDINAL
5,898,195 MONEY
1,362,396 MONEY
56 CARDINAL
23.1% PERCENT
6.5% PERCENT
10,000 MONEY
252 CARDINAL
1266 DATE
183,414 MONEY
79,329 MONEY
Technical Criminal Services ORG
788 $ MONEY
64,435 MONEY
First ORDINAL
755 $61,696 MONEY
687 CARDINAL
60,501 MONEY
Emergency Services ORG
Homeland Security ORG
Emergency Services ORG
692 $54,374 MONEY
First ORDINAL
635 $49,793 MONEY
49,311 MONEY
Emergency Services ORG
Homeland Security ORG
Civil Defense PERSON
812 $48,700 MONEY
47,619 MONEY
45,353 MONEY
Emergency Services ORG
Homeland Security ORG
Fire Safety PERSON
44,586 MONEY

493 $ MONEY
39,505 MONEY
31,041 MONEY
Personnel ORG
362 $ MONEY
30,087 MONEY
302 $ MONEY
24,241 MONEY
Public Works ORG
223 CARDINAL
23,570 MONEY
301 $ MONEY
22,277 MONEY
Emergency Services ORG
Homeland Security Special ORG
480 $ MONEY
21,446 MONEY
Personnel ORG
349 $ MONEY
20,836 MONEY
County Executive ORG
County Executive ORG
173 $ MONEY
20,506 MONEY
Emergency Services ORG
Homeland Security ORG
342 CARDINAL
20,480 MONEY
213 CARDINAL
18,337 MONEY
370 CARDINAL
17,508 MONEY
Labor Relations ORG
165 $ MONEY
12,857 MONEY
12,626 MONEY
189 CARDINAL
11,840 MONEY
more than $10,000 MONEY
14 CARDINAL
the County Executive ORG
13 CARDINAL
March DATE
2020 DATE
200 hours TIME
CARES Act Guidance Regarding Meals LAW
COVID-19 ORG
COVID-19 ORG
COVID-19 PERSON
252 CARDINAL

Erie County GPE
RECOMMEND ORG
the CARES Act LAW
New York GPE
The New York State Office of the Comptroller ORG
the Office of the New York State Comptroller ORG
County GPE
SAP GPE
9th ORDINAL
COVID-19 ORG
March DATE
New York GPE
RECOMMEND ORG
Appendix A. PERSON
December 31, 2020 DATE
252 CARDINAL
239,689 MONEY
Erie County GPE
Non-Food Items ORG
CARES Act LAW
New York State Comptroller ORG
CARES Act LAW
March 1, 2020 DATE
December 31, 2021 DATE
Health Department — Bioterrorism ORG
February 18, 2020 DATE
February 22, 2020 DATE
CARES Act LAW
Communal Food: Many ORG
The American Industrial Hygiene Association ORG
COVID-19 ORG
Possible Duplicate Payment WORK_OF_ART
Wellington Pub ORG
55 CARDINAL
the SH Hotel ORG
May 3, 2020 DATE
5716 DATE
05032020-0001 DATE
Wellington Pub ORG
55 CARDINAL
the SH Hotel ORG
May 3, 2020 DATE
5716 DATE
2020VF-119 CARDINAL
Duplicate Payment - ECC WORK_OF_ART
ECC Food Service ORG
72 CARDINAL
Swedish NORP
May 13, 2020 DATE
400 CARDINAL

ECC Food Service ORG
72 CARDINAL
Swedish NORP
May 13, 2020 DATE
400 CARDINAL
two CARDINAL
ECC Food Service ORG
days DATE
only one CARDINAL
ECC Food Service ORG
SAP ORG
SAP ORG
the Comptroller's Office ORG
November DATE
The CARES Act LAW
Erie County GPE
$160 million MONEY
COVID-19 ORG
COVID-19 ORG
The Comptroller's Office ORG
Legislature ORG
CARES Act PRODUCT
Legislature ORG
January 1, 2021 DATE
January 8, 2021 DATE
1,226 CARDINAL
114,524 hours TIME
252 CARDINAL
5,898,195 MONEY
1,362,396 MONEY
56 CARDINAL
183,414 MONEY
Erie County GPE
five CARDINAL
over 4,700 CARDINAL
five CARDINAL
202,312 MONEY
New York State GPE
Erie County GPE
Erie County's GPE
December 31, 2020 DATE
252 CARDINAL
239,689 MONEY
RECOMMEND ORG
CARES Act LAW
the CARES Act LAW
RECOMMEND ORG
Comm GPE
2020 DATE
RECOMMEND ORG

RECOMMEND ORG
RECOMMEND ORG
11 CARDINAL
252 CARDINAL
12 CARDINAL
2/18/2020 DATE
45.67 MONEY
3/16/2020 CARDINAL
2,618.01 MONEY
3/17/2020 CARDINAL
196.56 MONEY
3/19/20 $146.45 Food MONEY
Casa Di Pizza ORG
3/19/20 CARDINAL
146.96 MONEY
3/19/2020 CARDINAL
264.75 MONEY
3/19/20 $460.77 Food MONEY
Chef's ORG
3/20/2020 CARDINAL
100.80 MONEY
W B MASON CO INC ORG
3/20/2020 CARDINAL
239.75 MONEY
3/20/2020 DATE
264.75 MONEY
3/20/2020 $ MONEY
335.16 MONEY
W B MASON CO INC ORG
3/22/2020 DATE
134.88 MONEY
3/22/2020 DATE
304.71 MONEY
3/23/2020 $194.80 MONEY
3/23/2020 $840.00 MONEY
3/24/2020 DATE
822.50 MONEY
3/25/2020 DATE
717.50 MONEY
3/26/20 $53.92 Groceries MONEY
3/26/20 $272.34 Food MONEY
Casa Di Pizza ORG
3/26/2020 CARDINAL
708.75 MONEY
3/27/20 $27.19 Food MONEY
3/27/20 $51.46 MONEY
3/27/20 $53.15 Food MONEY
3/27/20 $88.86 Food MONEY
3/27/20 $99.11 Food MONEY
Picasso WORK_OF_ART

3/27/20 $444.57 Food MONEY
Chick N Pizza Works ORG
3/27/20 CARDINAL
534.60 MONEY
Chick N Pizza Works ORG
3/27/20 $634.38 Food MONEY
Chick N Pizza Works ORG
3/27/20 $654.36 Food MONEY
Chick N Pizza Works ORG
3/27/20 $680.26 Food MONEY
Chick N Pizza Works ORG
3/27/2020 CARDINAL
817.00 MONEY
3/28/2020 CARDINAL
214.80 MONEY
3/28/2020 DATE
257.73 MONEY
3/28/2020 DATE
284.73 MONEY
3/29/2020 DATE
264.75 MONEY
3/29/2020 DATE
264.75 MONEY
3/30/2020 CARDINAL
214.80 MONEY
3/30/2020 CARDINAL
294.72 MONEY
3/30/2020 $902.50 MONEY
3/31/2020 $807.50 MONEY
4/1/2020 DATE
807.50 MONEY
4/2/20 $82.12 Food Reimbursement MONEY
4/2/20 $419.55 Food MONEY
Chick N Pizza Works ORG
13 CARDINAL
4/2/20 CARDINAL
464.55 MONEY
Chick N Pizza Works ORG
4/2/2020 GPE
470.75 MONEY
4/2/2020 GPE
483.50 MONEY
4/2/20 $564.45 Food MONEY
Chick N Pizza Works ORG
4/2/20 $564.45 Food MONEY
Chick N Pizza Works ORG
4/3/2020 CARDINAL
608.00 MONEY
4/4/2020 $ MONEY
164.85 MONEY

4/4/2020 DATE
214.80 MONEY
4/5/2020 DATE
164.85 MONEY
4/5/2020 DATE
164.85 MONEY
4/5/2020 DATE
214.80 MONEY
4/5/2020 DATE
464.55 MONEY
4/5/2020 DATE
464.55 MONEY
4/6/2020 DATE
174.84 MONEY
CHICK N PIZZA PARTY WORKS ORG
4/6/2020 DATE
214.80 MONEY
4/6/2020 DATE
731.50 MONEY
4/6/2020 DATE
1,440.50 MONEY
4/7/2020 $ MONEY
174.84 MONEY
CHICK N PIZZA PARTY WORKS ORG
4/7/2020 $ MONEY
214.80 MONEY
4/7/2020 $ MONEY
627.00 MONEY
4/8/2020 $ MONEY
144.87 MONEY
4/8/2020 DATE
214.80 MONEY
4/8/2020 DATE
722.00 MONEY
4/9/20 $45.67 MONEY
Wegman PERSON
4/9/2020 DATE
154.86 MONEY
4/9/2020 DATE
214.80 MONEY
4/9/2020 DATE
655.50 MONEY
4/10/20 $38.31 Reimbursement MONEY
4/10/20 $51.99 Food MONEY
4/10/20 $56.12 Reimbursement MONEY
4/10/20 $63.77 Food MONEY
4/10/20 $65.24 Food MONEY
4/10/20 $68.08 Food MONEY
John PERSON
4/10/20 $71.25 MONEY

4/10/20 CARDINAL
4/10/20 $83.74 Food MONEY
4/10/20 $84.95 Food MONEY
Marotto GPE
4/10/20 $85.55 MONEY
Bocce Club Pizza ORG
4/10/20 $122.71 Food MONEY
Marotto GPE
4/10/20 $134.00 Food MONEY
Tony PERSON
4/10/2020 DATE
164.85 MONEY
4/10/20 $210.95 Food MONEY
La Nova ORG
4/10/2020 DATE
214.80 MONEY
4/10/20 $233.90 Food MONEY
Tony PERSON
14 CARDINAL
4/10/20 $277.34 Food MONEY
Casa Di PERSON
4/10/20 $431.50 Food MONEY
Tony PERSON
4/10/20 $526.70 Food MONEY
Tony PERSON
4/10/20 $612.00 Food MONEY
Tony PERSON
4/10/20 $648.00 Food MONEY
Poppyseed ORG
4/10/20 $680.00 Food MONEY
Poppyseed GPE
4/10/2020 DATE
694.32 MONEY
4/10/2020 DATE
712.50 MONEY
4/11/2020 CARDINAL
4/11/2020 $164.85 MONEY
4/11/2020 $164.85 MONEY
4/11/2020 CARDINAL
444.57 MONEY
4/11/2020 $608.00 MONEY
4/13/2020 $ MONEY
174.84 MONEY
CHICK N PIZZA PARTY WORKS ORG
4/13/2020 $ MONEY
214.80 MONEY
4/13/2020 $ MONEY
214.80 MONEY
4/13/2020 $589.00 MONEY
4/14/2020 $204.81 MONEY

4/14/2020 $494.00 MONEY
4/15/2020 $88.83 MONEY
CHICK N PIZZA PARTY WORKS ORG
4/15/2020 $164.85 MONEY
4/15/2020 DATE
541.50 MONEY
4/16/20 CARDINAL
196.56 MONEY
Wegman PERSON
4/16/2020 $ MONEY
214.80 MONEY
4/16/2020 $722.00 MONEY
4/16/20 CARDINAL
2,618.01 MONEY
Wegman PERSON
4/17/20 CARDINAL
40.96 MONEY
Response Team ORG
4/17/20 $ MONEY
76.11 MONEY
Response Team ORG
4/17/20 CARDINAL
76.66 MONEY
Response Team ORG
4/17/20 CARDINAL
114.90 MONEY
Response Team ORG
4/17/2020 CARDINAL
134.88 MONEY
4/17/20 $ MONEY
140.86 MONEY
Response Team ORG
4/17/20 $154.86 MONEY
Response Team ORG
4/17/20 $154.86 MONEY
Response Team ORG
4/17/2020 CARDINAL
174.84 MONEY
CHICK N PIZZA PARTY WORKS ORG
4/17/2020 DATE
214.80 MONEY
4/17/20 CARDINAL
214.80 MONEY
Response Team ORG
4/17/20 CARDINAL
214.80 MONEY
Response Team ORG
4/17/20 CARDINAL
214.80 MONEY
Response Team ORG

4/17/20 CARDINAL
304.71 MONEY
Response Team ORG
4/17/20 CARDINAL
347.63 MONEY
Response Team ORG
4/17/20 CARDINAL
491.00 MONEY
Response Team ORG
4/17/20 $ MONEY
532.65 MONEY
Response Team ORG
4/17/20 CARDINAL
544.55 MONEY
Response Team ORG
4/17/2020 CARDINAL
579.50 MONEY
15 CARDINAL
4/17/20 CARDINAL
664.00 MONEY
Response Team ORG
4/17/20 CARDINAL
704.00 MONEY
Response Team ORG
4/18/2020 $204.81 MONEY
4/18/2020 $204.81 MONEY
4/19/2020 DATE
244.77 MONEY
4/19/2020 DATE
244.77 MONEY
4/20/2020 $214.80 MONEY
4/20/2020 $234.78 MONEY
4/20/2020 $264.75 MONEY
4/20/2020 $655.50 MONEY
4/21/2020 DATE
1,180.25 MONEY
4/22/2020 $234.78 MONEY
4/22/2020 DATE
674.50 MONEY
4/23/20 $47.33 Reimbursement MONEY
4/23/20 $68.08 Food MONEY
John PERSON
4/23/20 $84.38 Food MONEY
4/23/20 $85.48 Food MONEY
John PERSON
4/23/20 $198.77 Food MONEY
Wellington Pub ORG
4/23/20 $214.75 Food MONEY
Wellington Pub ORG
4/23/20 $302.64 Food MONEY

Wellington Pub ORG
4/23/20 $414.50 Food MONEY
Wellington Pub ORG
4/23/20 $437.45 Food MONEY
Tony PERSON
4/23/20 $688.00 Food MONEY
Poppyseed ORG
4/23/20 $688.00 Food MONEY
Poppyseed ORG
4/23/20 $695.13 Food MONEY
Wellington Pub ORG
4/23/2020 DATE
1,155.75 MONEY
4/24/2020 $44.99 MONEY
4/24/2020 $224.80 MONEY
4/24/2020 $579.50 MONEY
4/25/2020 GPE
264.75 MONEY
4/26/2020 $ MONEY
264.75 MONEY
4/27/2020 DATE
234.78 MONEY
4/27/2020 DATE
264.75 MONEY
4/28/2020 DATE
214.80 MONEY
4/29/2020 $64.95 MONEY
CHICK N PIZZA PARTY WORKS ORG
4/30/2020 $655.50 MONEY
4/30/2020 $1,453.75 MONEY
5/1/2020 $ MONEY
224.80 MONEY
5/1/2020 DATE
769.50 MONEY
5/2/2020 CARDINAL
554.46 MONEY
5/3/2020 DATE
554.46 MONEY
5/4/2020 $154.86 MONEY
5/4/2020 $534.48 MONEY
5/4/2020 $ MONEY
893.00 MONEY
5/4/2020 $ MONEY
1,702.10 MONEY
16 CARDINAL
214.80 MONEY
5/8/2020 $ MONEY
94.92 MONEY
5/8/2020 $164.85 MONEY
5/8/20 $175.00 Food MONEY

Poppyseed GPE
5/8/20 $175.00 Food MONEY
Poppyseed GPE
5/8/20 $190.78 Food MONEY
Wellington Pub GPE
196.90 MONEY
214.75 MONEY
Wellington Pub GPE
214.75 MONEY
Wellington Pub GPE
214.80 MONEY
5/8/20 GPE
232.23 MONEY
Wellington Pub GPE
304.30 MONEY
Wellington Pub GPE
664.00 MONEY
Poppyseed GPE
5/8/20 $680.20 Food MONEY
704.00 MONEY
5/8/2020 $ MONEY
798.00 MONEY
5/9/2020 DATE
464.55 MONEY
5/10/2020 DATE
344.67 MONEY
5/10/2020 DATE
474.54 MONEY
5/11/2020 DATE
104.98 MONEY
5/11/2020 DATE
464.55 MONEY
5/11/2020 DATE
696.00 MONEY
5/11/2020 DATE
696.00 MONEY
5/11/2020 $ MONEY
779.00 MONEY
5/12/2020 $255.00 MONEY
5/12/2020 $255.00 MONEY
5/12/2020 $703.00 MONEY
5/13/20 $118.87 Food MONEY
Wellington Pub ORG
5/13/20 CARDINAL
175.00 MONEY
5/13/20 CARDINAL
175.00 MONEY
5/13/20 $190.31 Food MONEY
5/13/20 CARDINAL
190.78 MONEY

Wellington Pub ORG
5/13/20 CARDINAL
190.78 MONEY
Wellington Pub ORG
5/13/20 CARDINAL
196.90 MONEY
5/13/20 $196.90 Food MONEY
5/13/20 $196.90 Food MONEY
5/13/20 $196.90 Food MONEY
5/13/2020 $ MONEY
214.80 MONEY
5/13/20 CARDINAL
255.00 MONEY
5/13/20 $302.64 Food MONEY
Wellington Pub ORG
5/13/20 CARDINAL
402.75 MONEY
5/13/20 $430.48 Food MONEY
Wellington Pub ORG
5/13/20 CARDINAL
430.48 MONEY
Wellington Pub ORG
5/13/20 $454.45 Food MONEY
Wellington Pub ORG
17 CARDINAL
Wellington Pub ORG
5/13/20 $454.45 Food MONEY
Wellington Pub ORG
5/13/20 CARDINAL
534.35 MONEY
Wellington Pub GPE
5/13/2020 CARDINAL
640.00 MONEY
5/13/2020 $ MONEY
640.00 MONEY
5/13/20 $680.00 Food MONEY
Poppyseed GPE
5/13/2020 CARDINAL
684.00 MONEY
5/13/2020 $ MONEY
684.00 MONEY
5/13/20 $687.00 Food MONEY
5/13/20 CARDINAL
704.00 MONEY
5/13/20 $751.80 Food MONEY
5/13/20 $760.75 Food MONEY
5/14/2020 DATE
214.80 MONEY
5/14/20 CARDINAL
470.75 MONEY

Wegman PERSON
5/14/2020 CARDINAL
712.50 MONEY
5/14/20 CARDINAL
1,440.50 MONEY
Wegman PERSON
5/15/2020 CARDINAL
151.00 MONEY
5/15/2020 $151.00 MONEY
5/15/2020 DATE
655.50 MONEY
5/16/2020 CARDINAL
184.83 MONEY
5/16/2020 CARDINAL
404.61 MONEY
5/17/2020 $ MONEY
404.61 MONEY
5/18/2020 DATE
184.83 MONEY
5/18/2020 DATE
532.00 MONEY
5/18/2020 DATE
768.00 MONEY
5/19/2020 DATE
ASC FOOD SERVICES ORG
5/19/2020 DATE
ASC FOOD SERVICES ORG
5/19/2020 DATE
255.00 MONEY
5/19/2020 DATE
289.00 MONEY
5/19/2020 DATE
300.00 MONEY
5/19/2020 DATE
589.00 MONEY
5/20/20 $37.92 Reimbursement MONEY
5/20/2020 DATE
144.86 MONEY
5/20/20 CARDINAL
211.89 Cooking PERSON
5/20/20 CARDINAL
234.23 MONEY
5/20/2020 DATE
484.50 MONEY
5/20/2020 $927.00 MONEY
5/21/2020 DATE
214.80 MONEY
5/21/2020 DATE
342.00 MONEY
5/22/20 $77.97 Communication Boards MONEY

5/22/20 CARDINAL
199.90 MONEY
2020 DATE
Democrats NORP
Tonight TIME
Wisconsin GPE
19,500 CARDINAL
Trump ORG
Wisconsin GPE
Republicans NORP
Democrats NORP
Republicans NORP
Democrats NORP
Democrats NORP
Michigan AG ORG
Dana Nessel PERSON
Desist Order PERSON
first ORDINAL
Michigan GPE
6,000 CARDINAL
Trump PERSON
Oakland County GPE
Michigan GPE
Republican NORP
Pennsylvania GPE
nearly 20,000 CARDINAL
Trump PERSON
Biden PERSON
Biden PERSON
40,000 CARDINAL
Pennsylvania GPE
approximately 6,000 CARDINAL
Trump ORG
Biden PERSON
Georgia GPE
tonight TIME
approximately 10,000 CARDINAL
Trump PERSON
Biden PERSON
Wisconsin County GPE
Rock County Wisconsin GPE
election night TIME
the evening TIME
Trump PERSON
Wisconsin GPE
2016 DATE
10:59 TIME
around 29,000 CARDINAL
11:12 CARDINAL
Trump PERSON

nearly 1,000 CARDINAL
Biden PERSON
31,000 CARDINAL
Biden PERSON
30,000 CARDINAL
11:21 TIME
11:43 TIME
Trump ORG
46,649 CARDINAL
Biden ORG
37,133 CARDINAL
9,516 CARDINAL
Trump ORG
Biden PERSON
46,649 CARDINAL
Trump ORG
Biden PERSON
37,133 CARDINAL
Trump PERSON
Biden PERSON
Republican NORP
Democrat NORP
19,032 CARDINAL
tonight TIME
Biden PERSON
Trump PERSON
9,516 CARDINAL
Wisconsin GPE
Biden PERSON
1,630,570 CARDINAL
Trump PERSON
1,610,030 CARDINAL
Biden PERSON
1,508 CARDINAL
hundreds of thousands CARDINAL
Biden PERSON
Milwaukee GPE
early in the morning TIME
only one CARDINAL
Democrats NORP
Kim PERSON
2021 Act QUANTITY
Climate ORG
Dawn Euer PERSON
Lauren H. Carson PERSON
Rhode Island's GPE
House ORG
K. Joseph Shekarchi PERSON
Senate ORG
Dominick J. Ruggerio PERSON

```
today DATE
Dan McKee PERSON
Peter Neronha PERSON
Treasurer Seth Magaziner PERSON

dftrain["text"] = dftrain["text"].astype(str)

# Appliquer la fonction à la colonne "text"
dftrain["text"] = dftrain["text"].apply(add_entity_name)

print(dftrain["text"])

847      Also on the naughty list were Christmas (DATE)...
447      The number of people applying for UK (GPE) uni...
2243     Unvaccinated Children Pose No Risk To Anyone ,...
1202     Australia (GPE) 's Great (WORK_OF_ART) Barrier...
1059     One (CARDINAL) in ten (CARDINAL) patients admi...
                              ...
637      As Democrats (NORP) basked in the glory of the...
1022     HON (ORG) . STEFAN (PERSON) I. (PERSON) MYCHAJ...
1515     We 've reported on numerous events identified ...
1432     The 2021 (QUANTITY) Act (QUANTITY) on Climate ...
628      Antibody tests support what 's been obvious : ...
Name: text, Length: 1900, dtype: object
```

On scinde notre jeu de données en jeu de test et jeux d'apprentissage

```
X=dftrain['text']
X_train,X_test,y_train,y_test=train_test_split(X,y,test_size =
0.2,random_state=8)
print("X_train is \n",X_train.shape)
print("y_train is \n",y_train.shape)
print("X_test is \n",X_test.shape)
print("y_test is \n",y_test.shape)

X_train is
 (1520,)
y_train is
 (1520,)
X_test is
 (380,)
y_test is
 (380,)
```

Ici, c'est une étape importante, on va tester différents classifieurs, pour chacun des classifieurs, on va appliquer le prétraitement + Vectorisation TfIdf, et on applique une cross_val_score avec un Kfold de 10 fois, par la suite on stocke dans une liste all_results la moyenne des accuracy + l'écart type et on la trie par ordre décroissant de moyenne d'accuracy et d'écart type. on remarque que les 2 meilleurs sont SVM et RF qu'on va séléctionner pour leur appliquer le GridSearch sur les paramètres des prétraitements + leurs hyperparamètres pour pouvoir choisir le meilleur.

```python
score = 'accuracy'
seed = 7
allresults = []
results = []
names = []


# Liste des modèles à tester
models = [
    ('MultinomialNB', MultinomialNB()),
    ('LogisticRegression', LogisticRegression(random_state=42))
]


models.append(('KNN', KNeighborsClassifier()))
models.append(('CART', DecisionTreeClassifier()))
models.append(('RF', RandomForestClassifier()))
models.append(('SVM', SVC()))


# Création d'un pipeline pour chaque modèle
pipelines = []
for name,model in models:
    pipeline = Pipeline([
        ('normalize', TextNormalizer()),
        ('tfidf', TfidfVectorizer()),
        (name,model)
    ])
    pipelines.append((name,pipeline))
    #pipeline.fit(X_train_text,y_train)
all_results=[]
scores=[]
for p in pipelines:
    print(p[1])
    # cross validation en 10 fois
    kfold = KFold(n_splits=10,random_state=seed,shuffle=True)

#    print ("Evaluation de ",p)
    start_time = time.time()
    # application de la classification
    cv_results = cross_val_score(p[1],X_train,y_train, cv=kfold,
scoring=score)
    #print("Pour le classifieur",p[0],"on a un score
de",cv_results.mean(),"et un écart type de",cv_results.std())
    scores.append(cv_results)
    names.append(p[0])

    all_results.append((p[0],cv_results.mean(),cv_results.std()))
```

```python
    end_time = time.time()

all_results = sorted(all_results, key=lambda x: (-x[1], -x[2]))
print("all resultats", all_results)
    # affichage des résultats
#print ('\nLe meilleur resultat : ',max(results))
```

Pipeline(steps=[('normalize', TextNormalizer()), ('tfidf',
TfidfVectorizer()),
                ('MultinomialNB', MultinomialNB())])
Pipeline(steps=[('normalize', TextNormalizer()), ('tfidf',
TfidfVectorizer()),
                ('LogisticRegression',
LogisticRegression(random_state=42))])
Pipeline(steps=[('normalize', TextNormalizer()), ('tfidf',
TfidfVectorizer()),
                ('KNN', KNeighborsClassifier())])
Pipeline(steps=[('normalize', TextNormalizer()), ('tfidf',
TfidfVectorizer()),
                ('CART', DecisionTreeClassifier())])
Pipeline(steps=[('normalize', TextNormalizer()), ('tfidf',
TfidfVectorizer()),
                ('RF', RandomForestClassifier())])
Pipeline(steps=[('normalize', TextNormalizer()), ('tfidf',
TfidfVectorizer()),
                ('SVM', SVC())])
all resultats [('RF', 0.8710526315789474, 0.043857456085523605),
('SVM', 0.8697368421052631, 0.028646764548785834), ('CART',
0.8578947368421053, 0.031578947368421054), ('LogisticRegression',
0.8078947368421053, 0.03889933026077494), ('MultinomialNB',
0.7894736842105263, 0.03790752708128405), ('KNN', 0.6730263157894736,
0.030011828231465903)]

On affiche les boites à moustache pour mieux visualiser les résultats

```python
fig = plt.figure()
fig.suptitle('Comparaison des algorithmes')
ax = fig.add_subplot(111)
plt.boxplot(scores)
ax.set_xticklabels(names)
```
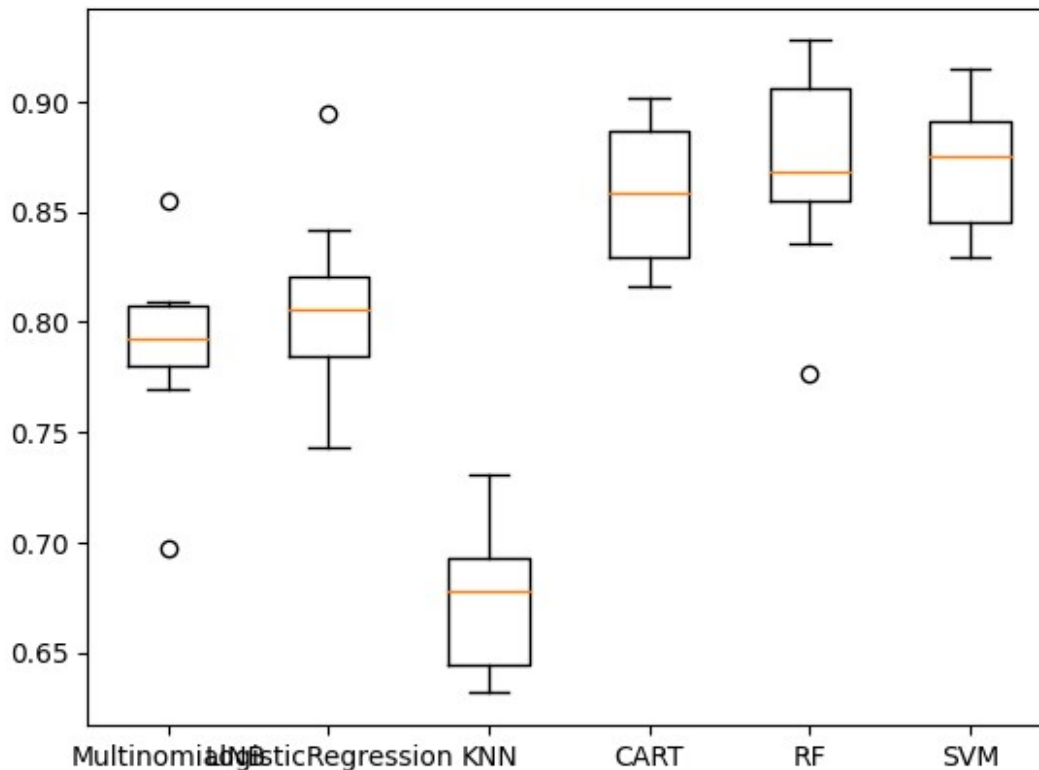
[Text(1, 0, 'MultinomialNB'),
 Text(2, 0, 'LogisticRegression'),
 Text(3, 0, 'KNN'),
 Text(4, 0, 'CART'),
 Text(5, 0, 'RF'),
 Text(6, 0, 'SVM')]

Comparaison des algorithmes

On a un pipeline pour chaque prétraitement différent, on essaye pas mal (miniscule, lemmatisation, miniscule + lemmatisation..) et on stocke le fit_transorm de nos X_train, X_test sur les pipelines dans des listes qui vont contenir tous les fit_transform des pipelines pour chaque classifieur, par la suite on parcourt ces listes là, on itère dessus, et chaque élement de la liste (train) va passer par le GridSearch et puis on predict sur son corresapondant dans liste (test).

```python
from sklearn.model_selection import GridSearchCV


from sklearn.datasets import fetch_20newsgroups
from sklearn.feature_extraction.text import CountVectorizer
from sklearn.feature_extraction.text import TfidfVectorizer
from sklearn.model_selection import train_test_split, GridSearchCV
from sklearn.pipeline import Pipeline
from sklearn.metrics import accuracy_score
from sklearn.naive_bayes import MultinomialNB
from sklearn.linear_model import LogisticRegression
from sklearn.svm import SVC
from sklearn.naive_bayes import MultinomialNB
from tabulate import tabulate
import numpy as np
np.random.seed(42) # Set the random seed for NumPy
```

```python
print("y_train", y_train.shape)
print("y_test", y_test.shape)
#X_test = np.ravel(X_test)
print("X_test", X_test.shape)
# le plus simple est de faire un test sur differents pipelines.


# pipeline de l'utilisation de TfidfVectorizer avec differents pre-
traitements
TFIDF_brut = Pipeline ([('cleaner', TextNormalizer()),
                        ('tfidf_vectorizer',
TfidfVectorizer(lowercase=False))])

TFIDF_lowcase = Pipeline([('cleaner',
TextNormalizer(removestopwords=False,lowercase=True,

getstemmer=False,removedigit=False)),
                        ('tfidf_vectorizer',
TfidfVectorizer(lowercase=False))])
TFIDF_lowStop = Pipeline([('cleaner',
TextNormalizer(removestopwords=True,lowercase=True,

getstemmer=False,removedigit=False)),
                        ('tfidf_vectorizer',
TfidfVectorizer(lowercase=False))])

TFIDF_lowStopstem = Pipeline([('cleaner',
TextNormalizer(removestopwords=True,lowercase=True,

getstemmer=True,removedigit=False)),
                        ('tfidf_vectorizer',
TfidfVectorizer(lowercase=False))])


# Liste de tous les modeles à tester
all_models = [
    ("TFIDF_lowcase", TFIDF_lowcase),
    ("TFIDF_lowStop", TFIDF_lowStop),
    ("TFIDF_lowStopstem",TFIDF_lowStopstem),
    ("TFIDF_brut", TFIDF_brut)
]

X_train_title_SVC = []
X_test_title_SVC = []




for name, pipeline in all_models :
```

```python
X_train_title_SVC.append(pipeline.fit_transform(X_train).toarray())
    X_test_title_SVC.append(pipeline.transform(X_test).toarray())



models = {
    'SVC': SVC(random_state=42)
}


params = {'SVC': [{'C': [0.001, 0.01, 0.1, 1,2]},
            {'gamma': [0.001, 0.01, 0.1,0.2,0.3]},
            {'kernel': ['linear', 'rbf']}]
}

for model_name, model in models.items():
    score='accuracy'
    X_train_title = eval('X_train_title_' + model_name)
    X_test_title = eval('X_test_title_' + model_name)
    for i in range (len(X_train_title)):
      grid_search = GridSearchCV(model, params[model_name], n_jobs=-1,
verbose=1,scoring=score)
      print("grid search fait")
      print("X_train",X_train_title[i].shape)
      print("y_train",y_train.shape)
      grid_search.fit(X_train_title[i],y_train)
      print ('meilleur score %0.3f'%(grid_search.best_score_),'\n')
      print ('meilleur estimateur',grid_search.best_estimator_,'\n')
      y_pred = grid_search.predict(X_test_title[i])
      MyshowAllScores(y_test,y_pred)

      print("Ensemble des meilleurs paramètres :")
      best_parameters = grid_search.best_estimator_.get_params()
      for param_dict in params[model_name]:
        for param_name, param_value in param_dict.items():
            print("\t%s: %r" % (param_name,
best_parameters[param_name]))
```

```
y_train (1520,)
y_test (380,)
X_test (380,)
grid search fait
X_train (1520, 29277)
y_train (1520,)
Fitting 5 folds for each of 12 candidates, totalling 60 fits
meilleur score 0.857

meilleur estimateur SVC(C=2, random_state=42)
```

```
Accuracy : 0.903
Classification Report
                precision    recall  f1-score   support

MIXTURE/OTHER     0.89385   0.89888   0.89636       178
   TRUE/FALSE     0.91045   0.90594   0.90819       202

     accuracy                         0.90263       380
    macro avg     0.90215   0.90241   0.90227       380
 weighted avg     0.90268   0.90263   0.90265       380


Ensemble des meilleurs paramètres :
     C: 2
     gamma: 'scale'
     kernel: 'rbf'
grid search fait
X_train (1520, 29142)
y_train (1520,)
Fitting 5 folds for each of 12 candidates, totalling 60 fits
meilleur score 0.857

meilleur estimateur SVC(C=2, random_state=42)

Accuracy : 0.905
Classification Report
                precision    recall  f1-score   support

MIXTURE/OTHER     0.89888   0.89888   0.89888       178
   TRUE/FALSE     0.91089   0.91089   0.91089       202

     accuracy                         0.90526       380
    macro avg     0.90488   0.90488   0.90488       380
 weighted avg     0.90526   0.90526   0.90526       380


Ensemble des meilleurs paramètres :
     C: 2
     gamma: 'scale'
     kernel: 'rbf'
grid search fait
X_train (1520, 19860)
y_train (1520,)
Fitting 5 folds for each of 12 candidates, totalling 60 fits
meilleur score 0.858

meilleur estimateur SVC(C=2, random_state=42)

Accuracy : 0.897
Classification Report
                precision    recall  f1-score   support
```

```
MIXTURE/OTHER     0.88398    0.89888    0.89136         178
   TRUE/FALSE     0.90955    0.89604    0.90274         202

     accuracy                           0.89737         380
    macro avg     0.89676    0.89746    0.89705         380
 weighted avg     0.89757    0.89737    0.89741         380


Ensemble des meilleurs paramètres :
     C: 2
     gamma: 'scale'
     kernel: 'rbf'
grid search fait
X_train (1520, 35329)
y_train (1520,)
Fitting 5 folds for each of 12 candidates, totalling 60 fits
meilleur score 0.863

meilleur estimateur SVC(C=2, random_state=42)

Accuracy : 0.897
Classification Report
             precision    recall   f1-score    support

MIXTURE/OTHER     0.89266    0.88764    0.89014         178
   TRUE/FALSE     0.90148    0.90594    0.90370         202

     accuracy                           0.89737         380
    macro avg     0.89707    0.89679    0.89692         380
 weighted avg     0.89735    0.89737    0.89735         380


Ensemble des meilleurs paramètres :
     C: 2
     gamma: 'scale'
     kernel: 'rbf'
```
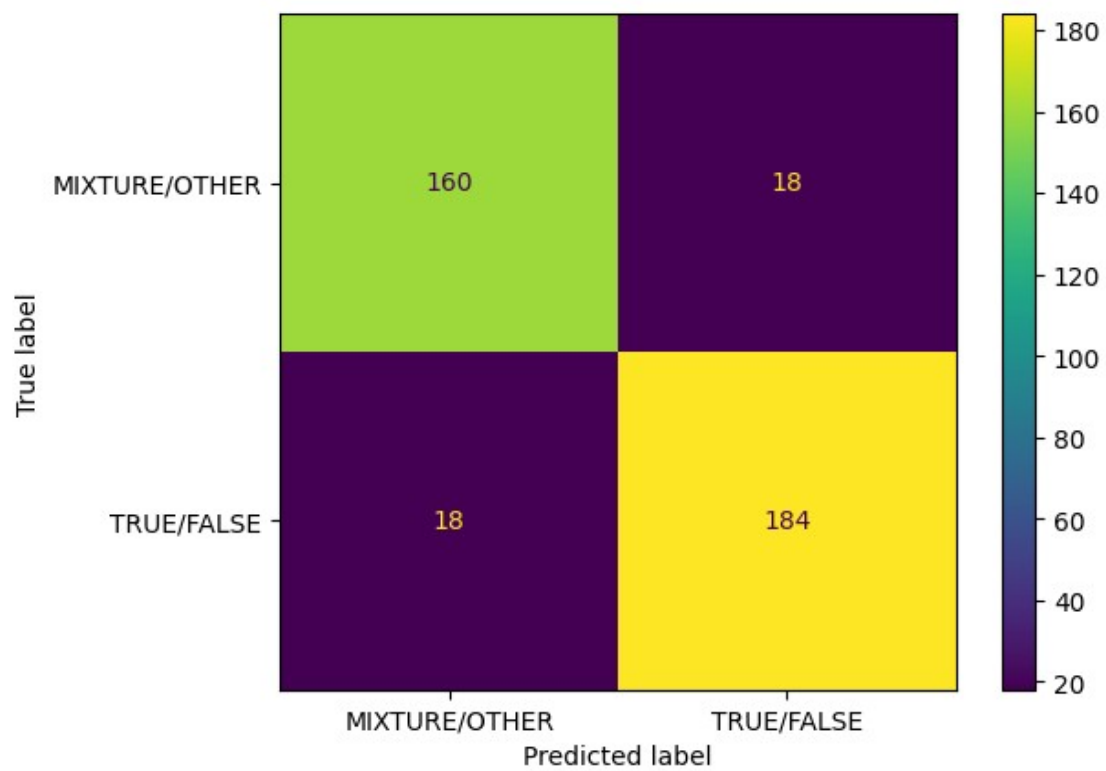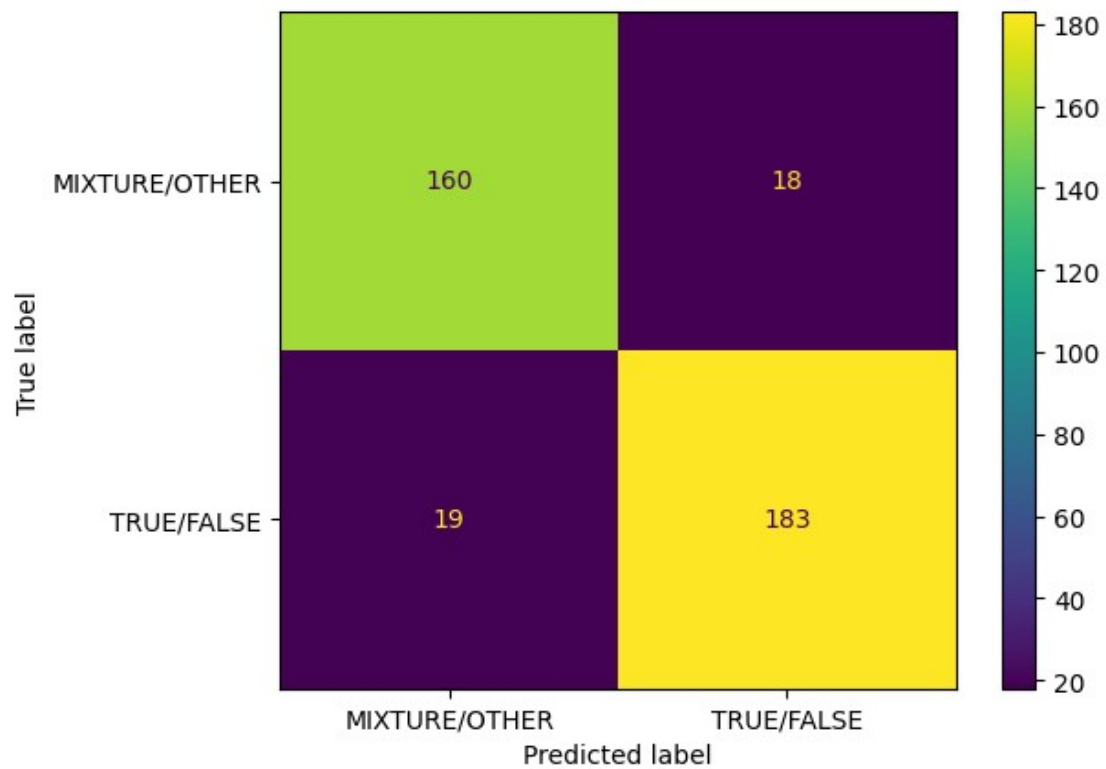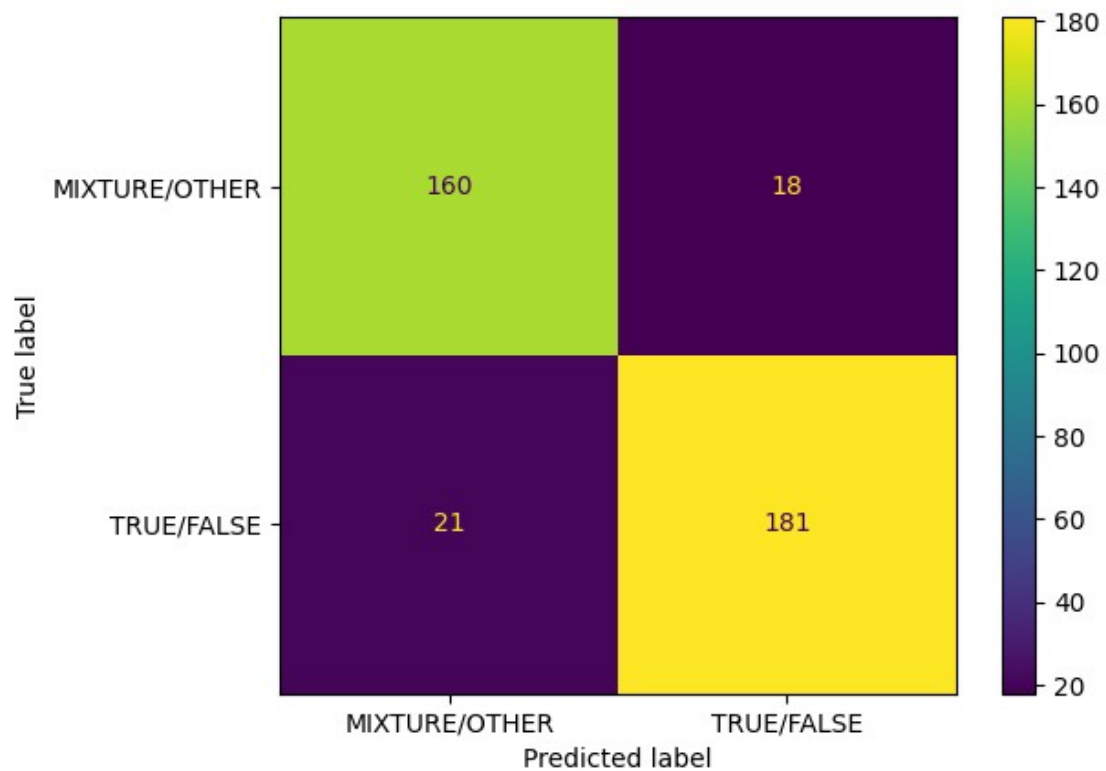
Etape 4 : Classification selon la colonne Text_title (concaténation de la colonne text et la colonne title) :

On concatène les deux colonnes text et titre de note DataFrame dftrain

```
text_title = dftrain.apply(lambda x : '{}
{}'.format(x['text'],x['title']),axis=1)
dftrain['text_title'] = text_title
X_text_title = dftrain.iloc[0:,-1]
X_text_title.reset_index(drop = True, inplace = True)
```

Ici, Pas besoin d'appliquer les entités entité dans cette étape vu qu'on les a appliqué sur les colonnes texte et title et ici on les concaténes donc on les a automatiquement

On scinde notre jeu de données en jeu d'apprentissage et jeu de test

```
X=dftrain['text_title']
X_train,X_test,y_train,y_test=train_test_split(X,y,test_size =
0.2,random_state=8)
print("X_train is \n",X_train.shape)
print("y_train is \n",y_train.shape)
print("X_test is \n",X_test.shape)
print("y_test is \n",y_test.shape)
```

Ici, c'est une étape importante, on va tester différents classifieurs, pour chacun des classifieurs, on va appliquer le prétraitement + Vectorisation TfIdf, et on applique une cross_val_score avec un Kfold de 10 fois, par la suite on stocke dans une liste all_results la moyenne des accuracy + l'écart type et on la trie par ordre décroissant de moyenne d'accuracy et d'écart type. on remarque que les 2 meilleurs sont SVM et RF qu'on va sélectionner pour leur appliquer le GridSearch sur les paramètres des prétraitements + leurs hyperparamètres pour pouvoir choisir le meilleur

```
score = 'accuracy'
seed = 7
allresults = []
results = []
names = []


# Liste des modèles à tester
models = [
    ('MultinomialNB', MultinomialNB()),
    ('LogisticRegression', LogisticRegression(random_state=42))
]


#models.append(('LR', LogisticRegression(solver='lbfgs')))
models.append(('KNN', KNeighborsClassifier()))
models.append(('CART', DecisionTreeClassifier(random_state=42)))
models.append(('RF', RandomForestClassifier(random_state=42)))
```

```python
models.append(('SVM', SVC(random_state=42)))

# Création d'un pipeline pour chaque modèle
pipelines = []
for name,model in models:
    pipeline = Pipeline([
        ('normalize', TextNormalizer()),
        ('tfidf', TfidfVectorizer()),
        (name,model)
    ])
    pipelines.append((name,pipeline))

all_results=[]
scores=[]
for p in pipelines:
    print(p[1])
    # cross validation en 10 fois
    kfold = KFold(n_splits=10,random_state=seed,shuffle=True)

#    print ("Evaluation de ",p)
    start_time = time.time()
    # application de la classification
    cv_results = cross_val_score(p[1],X_train,y_train, cv=kfold,
scoring=score)
    scores.append(cv_results)
    names.append(p[0])
    all_results.append((p[0],cv_results.mean(),cv_results.std()))
    end_time = time.time()


print("all resultats", all_results)

all_results = sorted(all_results, key=lambda x: (-x[1], -x[2]))
print("all resultats", all_results)
    # affichage des résultats
#print ('\nLe meilleur resultat : ',max(results))

Pipeline(steps=[('normalize', TextNormalizer()), ('tfidf',
TfidfVectorizer()),
                ('MultinomialNB', MultinomialNB())])
Pipeline(steps=[('normalize', TextNormalizer()), ('tfidf',
TfidfVectorizer()),
                ('LogisticRegression',
LogisticRegression(random_state=42))])
Pipeline(steps=[('normalize', TextNormalizer()), ('tfidf',
TfidfVectorizer()),
                ('KNN', KNeighborsClassifier())])
Pipeline(steps=[('normalize', TextNormalizer()), ('tfidf',
TfidfVectorizer()),
                ('CART', DecisionTreeClassifier(random_state=42))])
```

```
Pipeline(steps=[('normalize', TextNormalizer()), ('tfidf',
TfidfVectorizer()),
               ('RF', RandomForestClassifier(random_state=42))])
Pipeline(steps=[('normalize', TextNormalizer()), ('tfidf',
TfidfVectorizer()),
               ('SVM', SVC(random_state=42))])
all resultats [('MultinomialNB', 0.7861842105263157,
0.04464483572979879), ('LogisticRegression', 0.8144736842105263,
0.04291975376394759), ('KNN', 0.6848684210526316,
0.02889498849473387), ('CART', 0.8480263157894736,
0.032021298746415676), ('RF', 0.875657894736842, 0.03511580564495399),
('SVM', 0.8763157894736843, 0.0320685726498788)]
all resultats [('SVM', 0.8763157894736843, 0.0320685726498788), ('RF',
0.875657894736842, 0.03511580564495399), ('CART', 0.8480263157894736,
0.032021298746415676), ('LogisticRegression', 0.8144736842105263,
0.04291975376394759), ('MultinomialNB', 0.7861842105263157,
0.04464483572979879), ('KNN', 0.6848684210526316,
0.02889498849473387)]
```

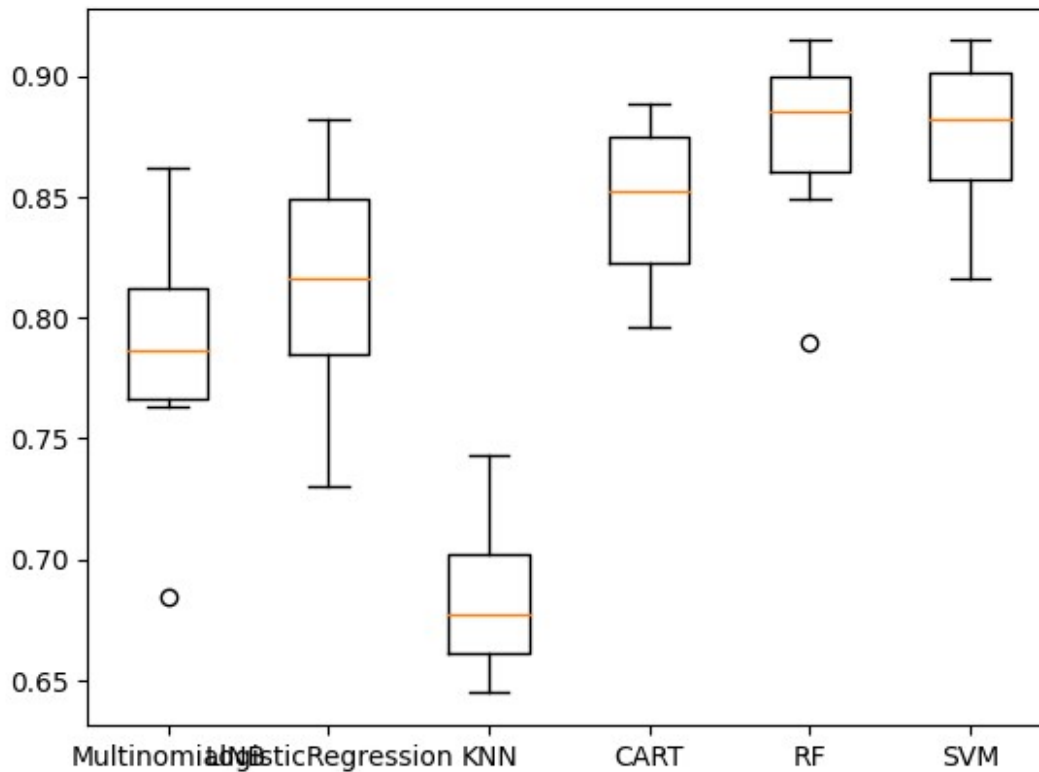On affiche les boites à moustache pour mieux visualiser les résultats

```
fig = plt.figure()
fig.suptitle('Comparaison des algorithmes')
ax = fig.add_subplot(111)
plt.boxplot(scores)
ax.set_xticklabels(names)

[Text(1, 0, 'MultinomialNB'),
 Text(2, 0, 'LogisticRegression'),
 Text(3, 0, 'KNN'),
 Text(4, 0, 'CART'),
 Text(5, 0, 'RF'),
 Text(6, 0, 'SVM')]
```

Comparaison des algorithmes

On a un pipeline pour chaque prétraitement différent, on essaye pas mal (miniscule, lemmatisation, miniscule + lemmatisation..) et on stocke le fit_transorm de nos X_train, X_test sur les pipelines dans des listes qui vont contenir tous les fit_transform des pipelines pour chaque classifieur, par la suite on parcourt ces listes là, on itère dessus, et chaque élement de la liste (train) va passer par le GridSearch et puis on predict sur son corresapondant dans liste (test).

```python
np.random.seed(42)  # Set the random seed for NumPy

#X_test = np.ravel(X_test)
# le plus simple est de faire un test sur differents pipelines.


# pipeline de l'utilisation de TfidfVectorizer avec differents pre-
traitements
TFIDF_brut = Pipeline ([('cleaner', TextNormalizer()),
                    ('tfidf_vectorizer',
TfidfVectorizer(lowercase=False))])

TFIDF_lowcase = Pipeline([('cleaner',
TextNormalizer(removestopwords=False,lowercase=True,
```

```python
                    getstemmer=False,removedigit=False)),
                                    ('tfidf_vectorizer',
TfidfVectorizer(lowercase=False))])
TFIDF_lowStop = Pipeline([('cleaner',
TextNormalizer(removestopwords=True,lowercase=True,

                    getstemmer=False,removedigit=False)),
                                    ('tfidf_vectorizer',
TfidfVectorizer(lowercase=False))])

TFIDF_lowStopstem = Pipeline([('cleaner',
TextNormalizer(removestopwords=True,lowercase=True,

                    getstemmer=True,removedigit=False)),
                                    ('tfidf_vectorizer',
TfidfVectorizer(lowercase=False))])


# Liste de tous les modeles à tester
all_models = [
    ("TFIDF_lowcase", TFIDF_lowcase),
    ("TFIDF_lowStop", TFIDF_lowStop),
    ("TFIDF_lowStopstem",TFIDF_lowStopstem),
    ("TFIDF_brut", TFIDF_brut)
]

X_train_title_SVC = []
X_test_title_SVC = []




for name, pipeline in all_models :

X_train_title_SVC.append(pipeline.fit_transform(X_train).toarray())
    X_test_title_SVC.append(pipeline.transform(X_test).toarray())



models = {
    'SVC': SVC(random_state=42)
}


params = {'SVC': [{'C': [0.001, 0.01, 0.1, 1,2]},
            {'gamma': [0.001, 0.01, 0.1,0.2,0.3]},
            {'kernel': ['linear', 'rbf']}]
}
```

```python
for model_name, model in models.items():
    score='accuracy'
    X_train_title = eval('X_train_title_' + model_name)
    X_test_title = eval('X_test_title_' + model_name)
    for i in range (len(X_train_title)):
        grid_search = GridSearchCV(model, params[model_name], n_jobs=-1,
verbose=1,scoring=score)
        print("grid search fait")
        print("X_train",X_train_title[i].shape)
        print("y_train",y_train.shape)
        grid_search.fit(X_train_title[i],y_train)
        print ('meilleur score %0.3f'%(grid_search.best_score_),'\n')
        print ('meilleur estimateur',grid_search.best_estimator_,'\n')
        y_pred = grid_search.predict(X_test_title[i])
        MyshowAllScores(y_test,y_pred)

        print("Ensemble des meilleurs paramètres :")
        best_parameters = grid_search.best_estimator_.get_params()
        for param_dict in params[model_name]:
            for param_name, param_value in param_dict.items():
                print("\t%s: %r" % (param_name,
best_parameters[param_name]))
```

```
grid search fait
X_train (1520, 29939)
y_train (1520,)
Fitting 5 folds for each of 12 candidates, totalling 60 fits
meilleur score 0.857

meilleur estimateur SVC(C=2, random_state=42)

Accuracy : 0.897
Classification Report
              precision    recall  f1-score   support

MIXTURE/OTHER   0.89266   0.88764   0.89014       178
   TRUE/FALSE   0.90148   0.90594   0.90370       202

    accuracy                       0.89737       380
   macro avg   0.89707   0.89679   0.89692       380
weighted avg   0.89735   0.89737   0.89735       380

Ensemble des meilleurs paramètres :
    C: 2
    gamma: 'scale'
    kernel: 'rbf'
grid search fait
X_train (1520, 29804)
y_train (1520,)
```

```
Fitting 5 folds for each of 12 candidates, totalling 60 fits
meilleur score 0.859

meilleur estimateur SVC(C=2, random_state=42)

Accuracy : 0.897
Classification Report
               precision    recall  f1-score   support

MIXTURE/OTHER    0.89266   0.88764   0.89014       178
  TRUE/FALSE     0.90148   0.90594   0.90370       202

    accuracy                         0.89737       380
   macro avg     0.89707   0.89679   0.89692       380
weighted avg     0.89735   0.89737   0.89735       380

Ensemble des meilleurs paramètres :
     C: 2
     gamma: 'scale'
     kernel: 'rbf'
grid search fait
X_train (1520, 20299)
y_train (1520,)
Fitting 5 folds for each of 12 candidates, totalling 60 fits
meilleur score 0.860

meilleur estimateur SVC(C=2, random_state=42)

Accuracy : 0.897
Classification Report
               precision    recall  f1-score   support

MIXTURE/OTHER    0.89266   0.88764   0.89014       178
  TRUE/FALSE     0.90148   0.90594   0.90370       202

    accuracy                         0.89737       380
   macro avg     0.89707   0.89679   0.89692       380
weighted avg     0.89735   0.89737   0.89735       380

Ensemble des meilleurs paramètres :
     C: 2
     gamma: 'scale'
     kernel: 'rbf'
grid search fait
X_train (1520, 36603)
y_train (1520,)
Fitting 5 folds for each of 12 candidates, totalling 60 fits
meilleur score 0.864
```

```
meilleur estimateur SVC(C=2, random_state=42)

Accuracy : 0.889
Classification Report
               precision    recall  f1-score   support

MIXTURE/OTHER    0.87778   0.88764   0.88268       178
   TRUE/FALSE    0.90000   0.89109   0.89552       202

     accuracy                        0.88947       380
    macro avg    0.88889   0.88936   0.88910       380
 weighted avg    0.88959   0.88947   0.88951       380

Ensemble des meilleurs paramètres :
     C: 2
     gamma: 'scale'
     kernel: 'rbf'
```
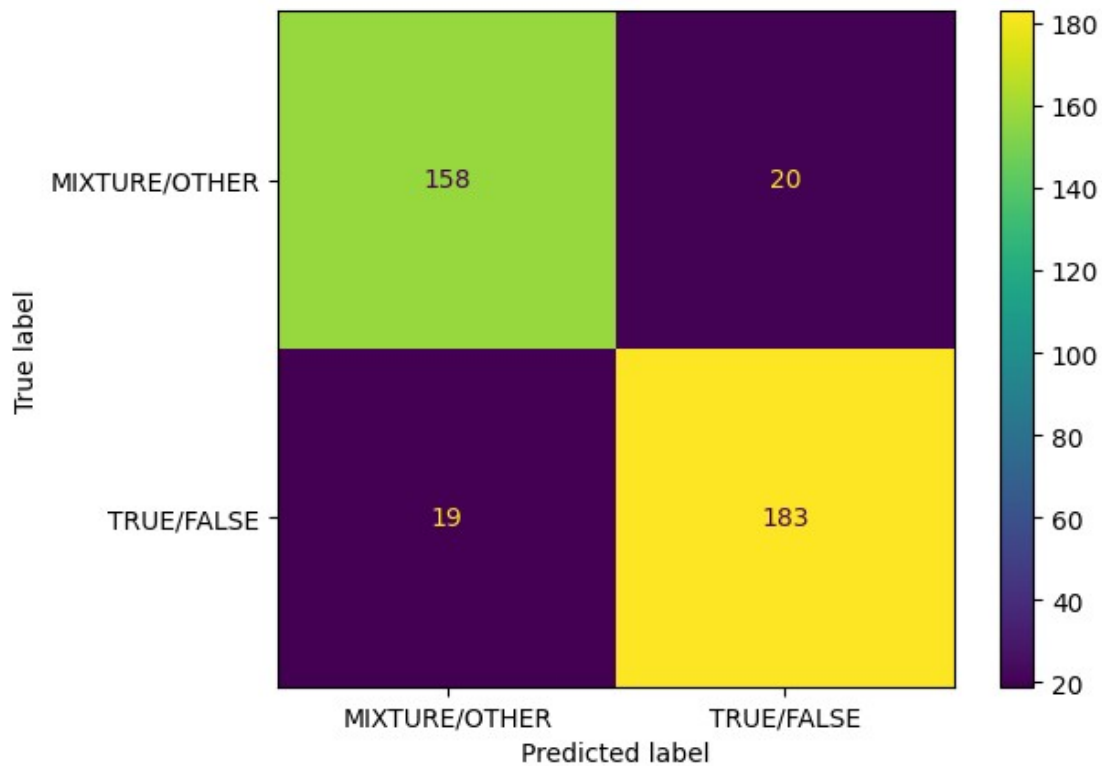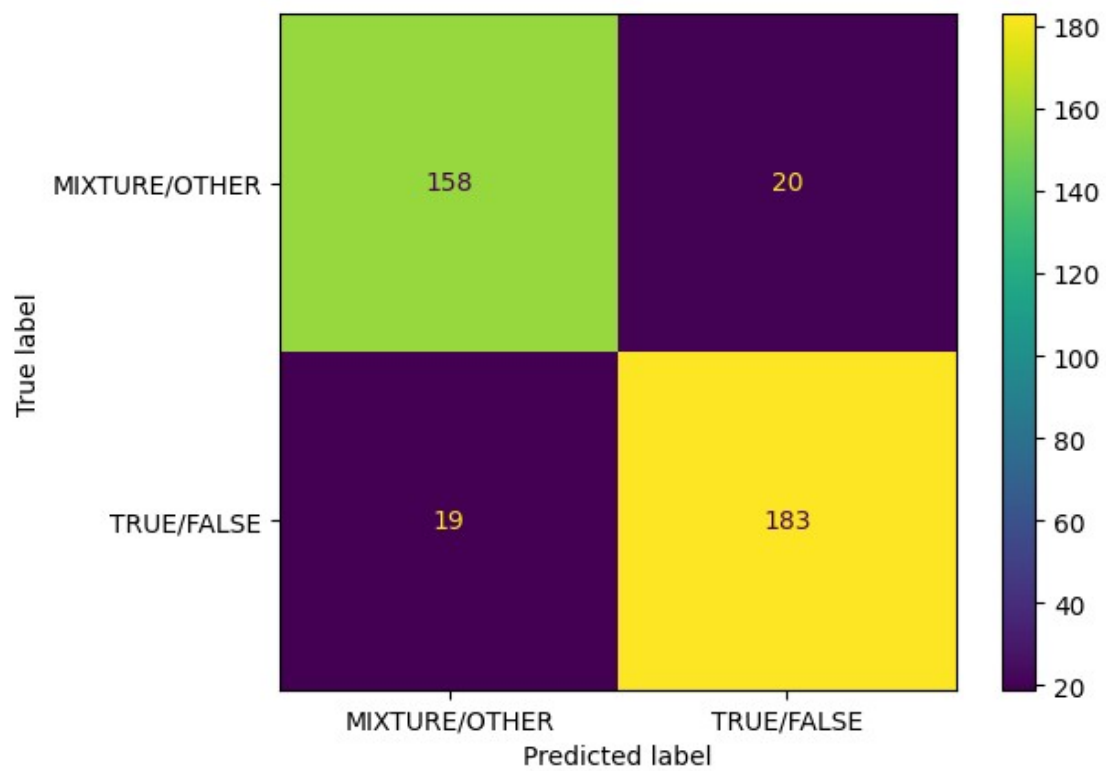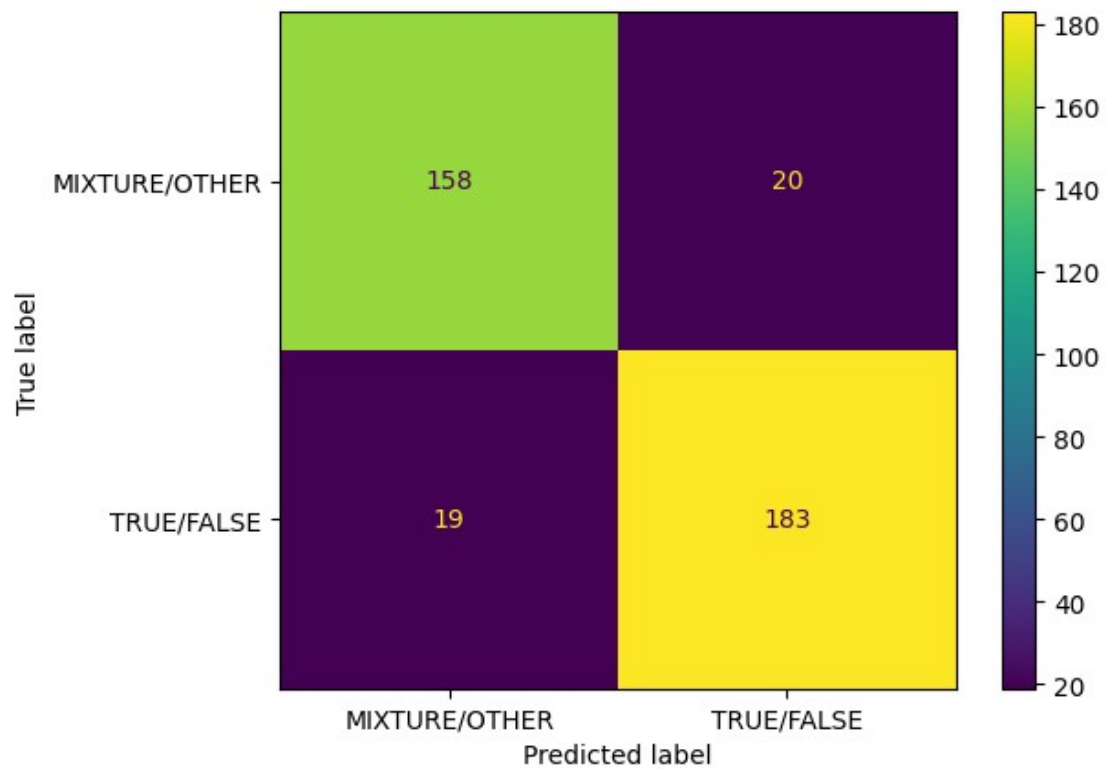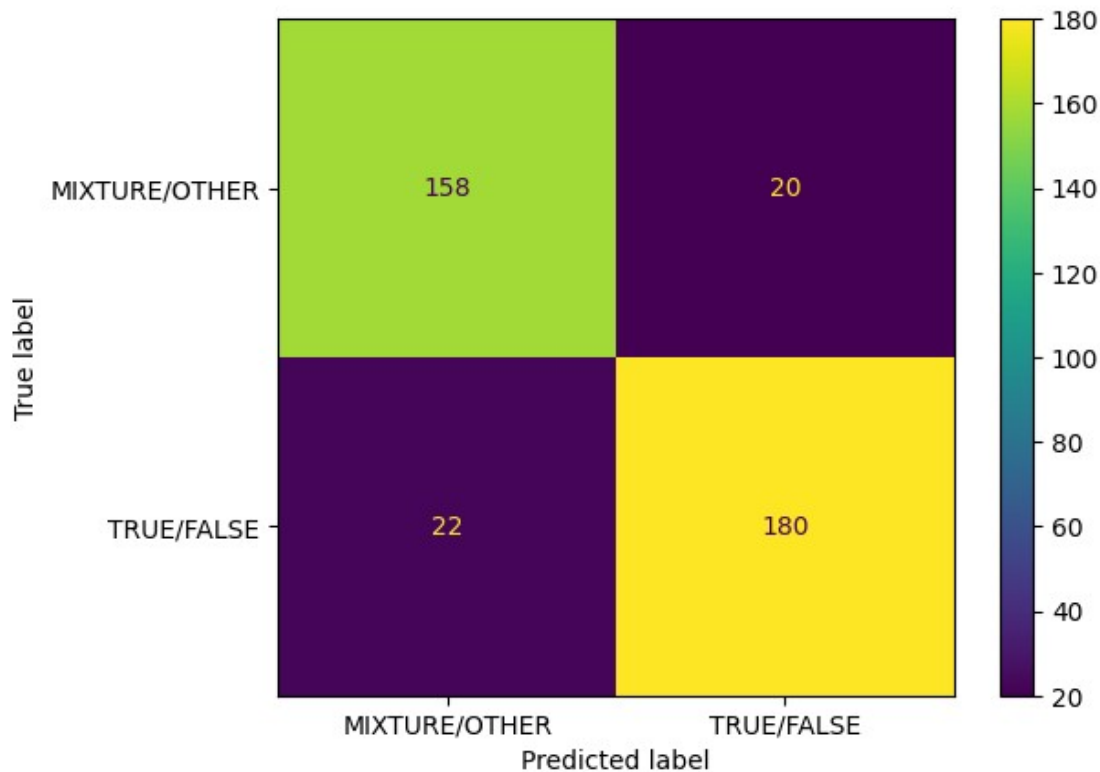
#Seconde classification: Other vs mixture

On crée un nouveau dataframe avec que les lignes contenant dans la colonne rating les valeurs other et mixture

```
dftrain = dftrainbase.loc[dftrainbase['rating'].isin(['other',
'mixture'])]
print(dftrain.shape)
print(dftrainbase.columns)
y = dftrain.iloc[:, -2]
print("les valeurs de OTHER et MIXTURE maintenant sont ",
y.value_counts())
```

```
(950, 5)
Index(['id', 'text', 'title', 'rating', 'regrouped'], dtype='object')
les valeurs de OTHER et MIXTURE maintenant sont  mixture    716
other       234
Name: rating, dtype: int64
```

Le jeu de données étant déséquilibré, on a pensé à appliquer le downsampling pour équilibrer nos données. on séléctionne des lignes aléatoirement de TRUE/FALSE de telle sorte que le nombre de lignes de TRUE/FALSE soit = au nbr de lignes de Other. et on mélange le DataFrame

```
# Compter le nombre d'observations dans chaque catégorie
mixture_count = dftrain['rating'].value_counts()['mixture']
other_count = dftrain['rating'].value_counts()['other']
```

```python
# Trouver le nombre minimum d'observations parmi les catégories
min_count = min( mixture_count, other_count)

# Sous-échantillonner les catégories pour équilibrer les quantités
mixture_sampled = dftrain[dftrain['rating'] ==
'mixture'].sample(min_count, random_state=42)
other_sampled = dftrain[dftrain['rating'] ==
'other'].sample(min_count, random_state=42)


# Concaténer les échantillons pour obtenir un nouveau dataframe
équilibré
dftrain = pd.concat([mixture_sampled,other_sampled])

# Mélanger aléatoirement les données
dftrain = dftrain.sample(frac=1, random_state=42)
print(dftrain)
X_text=dftrain["text"]
X_title=dftrain["title"]
y=dftrain.iloc[0:,-2]
print("\n")
print("la taille de y_train est " ,y.shape)
print("\n")
print("les valeurs de OTHER et MIXTURE maintenant sont
" ,y.value_counts())
```

```
              id                                               text  \
623     11aea5b0   By Jim Francesconi As we pass the governance t...
717     897917f0   Free schools are the 'modern engines of social...
2398    282ff13b   More than 11,000 criminal trials in England an...
676     347530a3   On Tuesday, radio show host John Fredricks sta...
231     adc0db84   Duration 0:47 Villaraigosa talks high-speed ra...
...          ...                                                ...
1217    21d1d3f2   RETRACTION, Feb. 21, 2019: This Op-Ed has been...
713     e22398b3   GETTY Leading economists are finally acknowled...
652     964e13e1   Iran Parliamentarians have moved a bill in ord...
889     905519d3   Diabetes prescriptions are costing the NHS in ...
773     c94b8dc2   Jeb Bush Really Is His Own Man By Stephanie Sc...

                                                  title   rating  \
623    DUP man dating Poots' daughter claims his 'tea...  mixture
717    Kremlin calls new U.S. sanctions illegal, says...  mixture
2398   Thousands of trials end over no-show witnesses...  mixture
676                         Warren Statement on Boeing     other
231    Antonio Villaraigosa rips Gavin Newsom over hi...  mixture
...                                                  ...      ...
1217   [RETRACTED] Media Hysteria: Climate Change 'He...  mixture
713    With Stickers, a Petition and Even a Middle Na...    other
652    Opportunities Foreclosed Equity Denied New Yor...    other
```

```
889    Renationalising railways supported by almost t...    other
773    Putin oversees hypersonic weapon test, says it...  mixture

            regrouped
623    MIXTURE/OTHER
717    MIXTURE/OTHER
2398   MIXTURE/OTHER
676    MIXTURE/OTHER
231    MIXTURE/OTHER
...              ...
1217   MIXTURE/OTHER
713    MIXTURE/OTHER
652    MIXTURE/OTHER
889    MIXTURE/OTHER
773    MIXTURE/OTHER

[468 rows x 5 columns]


la taille de y_train est  (468,)


les valeurs de OTHER et MIXTURE maintenant sont  mixture    234
other      234
Name: rating, dtype: int64
```

Etape 1 : Classification selon la colonne TITLE

On divise notre grand X en jeu de données d'apprentissage et de test (20% de test).

```python
X=dftrain["title"]
print(X)

X_train,X_test,y_train,y_test=train_test_split(X,y,test_size =
0.2,random_state=8)
print("X_train is",X_train)
print("y_train is",y_train)
print("X_test is",X_test)
print("y_test is",y_test)
```

```
0       You Can Be Fined $1,500 If Your Passenger Is U...
1         Missouri lawmakers condemn Las Vegas shooting
2       CBC Cuts Donald Trump's 'Home Alone 2' Cameo O...
3       Obama's Daughters Caught on Camera Burning US ...
4       Leaked Visitor Logs Reveal Schiff's 78 Visits ...
                            ...
2523    Taxman fails to answer four million calls a ye...
2524    Police catch 11-year-olds being used to sell d...
2525    Help to Buy Isa scandal: 500,000 first-time bu...
```

```
2526              A coke-snorting generation of hypocrites
2527              A coke-snorting generation of hypocrites
Name: title, Length: 2528, dtype: object
X_train is 497     Three-quarters of NHS maternity wards have no ...
2021    B.C. man with ALS chooses medically assisted d...
1032    Global warming data FAKED by government to fit...
1254    It's heartless to blame job losses on unemploy...
1734    Wealthy professionals most likely to drink reg...
                   ...
2181    Lisa Page Squeals: DNC Server Was Not Hacked B...
2409    Help to Buy Isa scandal: 500,000 first-time bu...
2033    Georgia Voters Sound the Alarm, Claim They're ...
1364    One in five dads have not seen their kids on F...
451     Charlie Crist disses idea of debate with Nan Rich
Name: title, Length: 2022, dtype: object
y_train is 497      mixture/other
2021       true/false
1032       true/false
1254    mixture/other
1734       true/false
              ...
2181    mixture/other
2409       true/false
2033    mixture/other
1364       true/false
451        true/false
Name: regrouped, Length: 2022, dtype: object
X_test is 1350     Is It True that Chief Justice John Roberts Vis...
1199    Jerry Brown Blames Climate Change for Californ...
1102    Britain's cocaine use doubles in last seven ye...
801     Labour reveals over 200,000 nurses have quit t...
158     Chelsea Clinton Ranked World's 18th Richest Woman
                   ...
1252            Climate Change Isn't the End of the World
1706    A look at Congressman Conor Lamb's voting reco...
265     Appalling: Oprah Says That When Done "Right" C...
1933            FG approves full reopening of schools
1421            A Letter from the Nye County Chairman
Name: title, Length: 506, dtype: object
y_test is 1350       true/false
1199    mixture/other
1102    mixture/other
801        true/false
158        true/false
              ...
1252    mixture/other
1706       true/false
265        true/false
1933       true/false
```

```
1421        true/false
Name: regrouped, Length: 506, dtype: object
```

Tester avec plusieurs classifieurs classiques.

```python
np.random.seed(42)   # Set the random seed for NumPy

score = 'accuracy'
seed = 7
allresults = []
results = []
names = []


# Liste des modèles à tester
models = [
    ('MultinomialNB', MultinomialNB()),
    ('LogisticRegression', LogisticRegression(random_state=42))
]


models.append(('KNN', KNeighborsClassifier()))
models.append(('CART', DecisionTreeClassifier(random_state=42)))
models.append(('RF', RandomForestClassifier(random_state=42)))
models.append(('SVM', SVC(random_state=42)))

# Création d'un pipeline pour chaque modèle
pipelines = []
for name,model in models:
    pipeline = Pipeline([
        ('normalize', TextNormalizer()),
        ('tfidf', TfidfVectorizer()),
        (name,model)
    ])
    pipelines.append((name,pipeline))
    #pipeline.fit(X_train,y_train)
all_results=[]
scores=[]
for p in pipelines:
    print(p[1])
    # cross validation en 10 fois
    kfold = KFold(n_splits=10,random_state=seed,shuffle=True)

#    print ("Evaluation de ",p)
    start_time = time.time()
    # application de la classification
    cv_results = cross_val_score(p[1],X_train,y_train, cv=kfold,
scoring=score)
    #print("Pour le classifieur",p[0],"on a un score
de",cv_results.mean(),"et un écart type de",cv_results.std())
```

```python
        scores.append(cv_results)
        names.append(p[0])

        all_results.append((p[0],cv_results.mean(),cv_results.std()))
        end_time = time.time()


print("all resultats", all_results)

all_results = sorted(all_results, key=lambda x: (-x[1], -x[2]))
print("all resultats", all_results)
```

```
X_train (374,)
y_train (374,)
Pipeline(steps=[('normalize', TextNormalizer()), ('tfidf',
TfidfVectorizer()),
                ('MultinomialNB', MultinomialNB())])
Pipeline(steps=[('normalize', TextNormalizer()), ('tfidf',
TfidfVectorizer()),
                ('LogisticRegression',
LogisticRegression(random_state=42))])
Pipeline(steps=[('normalize', TextNormalizer()), ('tfidf',
TfidfVectorizer()),
                ('KNN', KNeighborsClassifier())])
Pipeline(steps=[('normalize', TextNormalizer()), ('tfidf',
TfidfVectorizer()),
                ('CART', DecisionTreeClassifier(random_state=42))])
Pipeline(steps=[('normalize', TextNormalizer()), ('tfidf',
TfidfVectorizer()),
                ('RF', RandomForestClassifier(random_state=42))])
Pipeline(steps=[('normalize', TextNormalizer()), ('tfidf',
TfidfVectorizer()),
                ('SVM', SVC(random_state=42))])
all resultats [('MultinomialNB', 0.7995732574679943,
0.06954201329457245), ('LogisticRegression', 0.8291607396870555,
0.06914809365756731), ('KNN', 0.6608108108108108,
0.07455075868505182), ('CART', 0.8019203413940256,
0.04614638380816118), ('RF', 0.8663584637268847, 0.04660147638886548),
('SVM', 0.8800142247510669, 0.06314575928162124)]
all resultats [('SVM', 0.8800142247510669, 0.06314575928162124),
('RF', 0.8663584637268847, 0.04660147638886548),
('LogisticRegression', 0.8291607396870555, 0.06914809365756731),
('CART', 0.8019203413940256, 0.04614638380816118), ('MultinomialNB',
0.7995732574679943, 0.06954201329457245), ('KNN', 0.6608108108108108,
0.07455075868505182)]
```

Afifchage des boites à moustache pour mieux visualiser les résultats

```python
fig = plt.figure()
fig.suptitle('Comparaison des algorithmes')
```
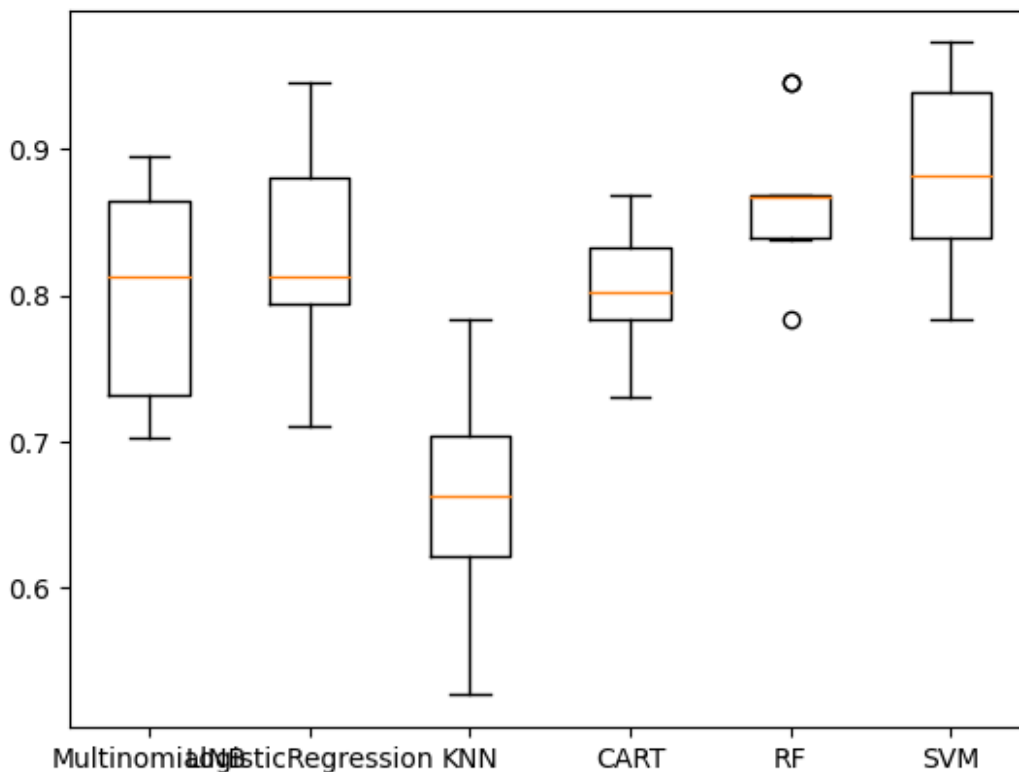
```
ax = fig.add_subplot(111)
plt.boxplot(scores)
ax.set_xticklabels(names)

[Text(1, 0, 'MultinomialNB'),
 Text(2, 0, 'LogisticRegression'),
 Text(3, 0, 'KNN'),
 Text(4, 0, 'CART'),
 Text(5, 0, 'RF'),
 Text(6, 0, 'SVM')]
```


Comparaison des algorithmes

Choisir les meilleurs paramètres et hyperparamètres pour SVM

```
np.random.seed(42)  # Set the random seed for NumPy
# pipeline de l'utilisation de TfidfVectorizer avec differents pre-
traitements
TFIDF_brut = Pipeline ([('cleaner', TextNormalizer()),
                  ('tfidf_vectorizer',
TfidfVectorizer(lowercase=False))])

TFIDF_lowcase = Pipeline([('cleaner',
TextNormalizer(removestopwords=False,lowercase=True,
```

```python
                            getstemmer=False,removedigit=False)),
                                    ('tfidf_vectorizer',
TfidfVectorizer(lowercase=False))])
TFIDF_lowStop = Pipeline([('cleaner',
TextNormalizer(removestopwords=True,lowercase=True,

                            getstemmer=False,removedigit=False)),
                                    ('tfidf_vectorizer',
TfidfVectorizer(lowercase=False))])

TFIDF_lowStopstem = Pipeline([('cleaner',
TextNormalizer(removestopwords=True,lowercase=True,

                            getstemmer=True,removedigit=False)),
                                    ('tfidf_vectorizer',
TfidfVectorizer(lowercase=False))])


# Liste de tous les modeles à tester
all_models = [
    ("TFIDF_lowcase", TFIDF_lowcase),
    ("TFIDF_lowStop", TFIDF_lowStop),
    ("TFIDF_lowStopstem",TFIDF_lowStopstem),
    ("TFIDF_brut", TFIDF_brut)
]

X_train_title_SVC = []
X_test_title_SVC = []



for name, pipeline in all_models :

X_train_title_SVC.append(pipeline.fit_transform(X_train).toarray())
    X_test_title_SVC.append(pipeline.transform(X_test).toarray())



models = {
    'SVC': SVC(random_state=42)
}


params = {'SVC': [{'C': [0.001, 0.01, 0.1, 1,2]},
            {'gamma': [0.001, 0.01, 0.1,0.2,0.3]},
            {'kernel': ['linear', 'rbf']}]
}
```

```python
for model_name, model in models.items():
    score='accuracy'
    X_train_title = eval('X_train_title_' + model_name)
    X_test_title = eval('X_test_title_' + model_name)
    for i in range (len(X_train_title)):
      grid_search = GridSearchCV(model, params[model_name], n_jobs=-1,
verbose=1,scoring=score)
      print("grid search fait")
      print("X_train",X_train_title[i].shape)
      print("y_train",y_train.shape)
      grid_search.fit(X_train_title[i],y_train)
      print ('meilleur score %0.3f'%(grid_search.best_score_),'\n')
      print ('meilleur estimateur',grid_search.best_estimator_,'\n')
      y_pred = grid_search.predict(X_test_title[i])
      MyshowAllScores(y_test,y_pred)

      print("Ensemble des meilleurs paramètres :")
      best_parameters = grid_search.best_estimator_.get_params()
      for param_dict in params[model_name]:
        for param_name, param_value in param_dict.items():
            print("\t%s: %r" % (param_name,
best_parameters[param_name]))
```

```
y_train (374,)
y_test (94,)
X_test (94,)
grid search fait
X_train (374, 2565)
y_train (374,)
Fitting 5 folds for each of 12 candidates, totalling 60 fits
meilleur score 0.858

meilleur estimateur SVC(C=1, random_state=42)

Accuracy : 0.926
Classification Report
              precision     recall  f1-score    support

     mixture    0.86667    0.97500   0.91765         40
       other    0.97959    0.88889   0.93204         54

    accuracy                         0.92553         94
   macro avg    0.92313    0.93194   0.92484         94
weighted avg    0.93154    0.92553   0.92591         94

Ensemble des meilleurs paramètres :
     C: 1
     gamma: 'scale'
     kernel: 'rbf'
grid search fait
```

```
X_train (374, 2445)
y_train (374,)
Fitting 5 folds for each of 12 candidates, totalling 60 fits
meilleur score 0.853

meilleur estimateur SVC(C=1, random_state=42)

Accuracy : 0.915
Classification Report
              precision    recall   f1-score    support

     mixture    0.86364    0.95000   0.90476         40
       other    0.96000    0.88889   0.92308         54

    accuracy                         0.91489         94
   macro avg    0.91182    0.91944   0.91392         94
weighted avg    0.91899    0.91489   0.91528         94

Ensemble des meilleurs paramètres :
     C: 1
     gamma: 'scale'
     kernel: 'rbf'
grid search fait
X_train (374, 2038)
y_train (374,)
Fitting 5 folds for each of 12 candidates, totalling 60 fits
meilleur score 0.853

meilleur estimateur SVC(C=1, random_state=42)

Accuracy : 0.926
Classification Report
              precision    recall   f1-score    support

     mixture    0.86667    0.97500   0.91765         40
       other    0.97959    0.88889   0.93204         54

    accuracy                         0.92553         94
   macro avg    0.92313    0.93194   0.92484         94
weighted avg    0.93154    0.92553   0.92591         94

Ensemble des meilleurs paramètres :
     C: 1
     gamma: 'scale'
     kernel: 'rbf'
grid search fait
X_train (374, 2914)
y_train (374,)
Fitting 5 folds for each of 12 candidates, totalling 60 fits
```
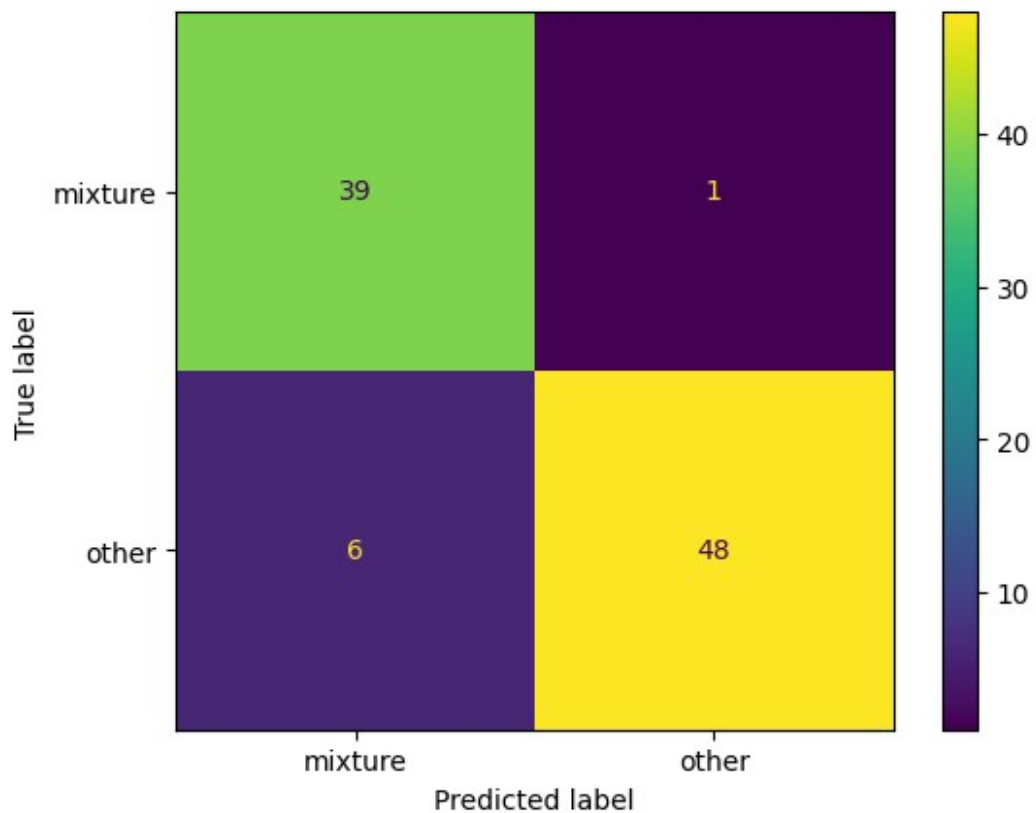
```
meilleur score 0.858

meilleur estimateur SVC(C=1, random_state=42)

Accuracy : 0.926
Classification Report
              precision    recall  f1-score   support

     mixture    0.86667   0.97500   0.91765        40
       other    0.97959   0.88889   0.93204        54

    accuracy                        0.92553        94
   macro avg    0.92313   0.93194   0.92484        94
weighted avg    0.93154   0.92553   0.92591        94

Ensemble des meilleurs paramètres :
    C: 1
    gamma: 'scale'
    kernel: 'rbf'
```
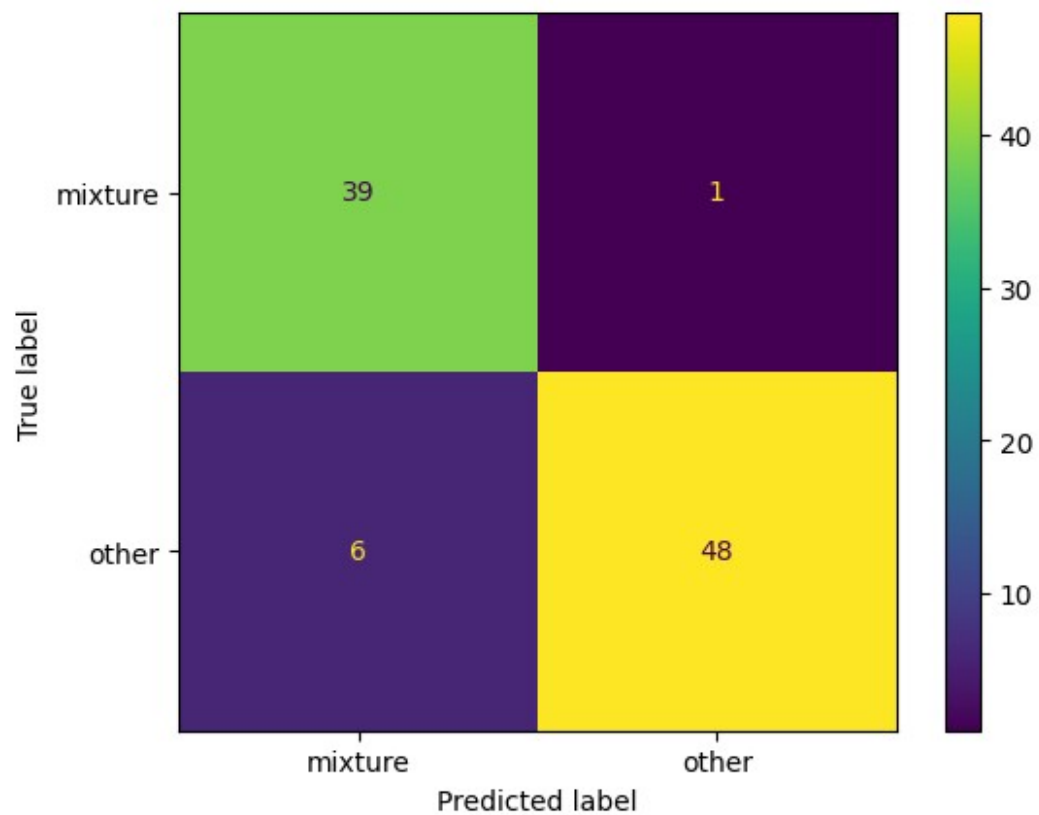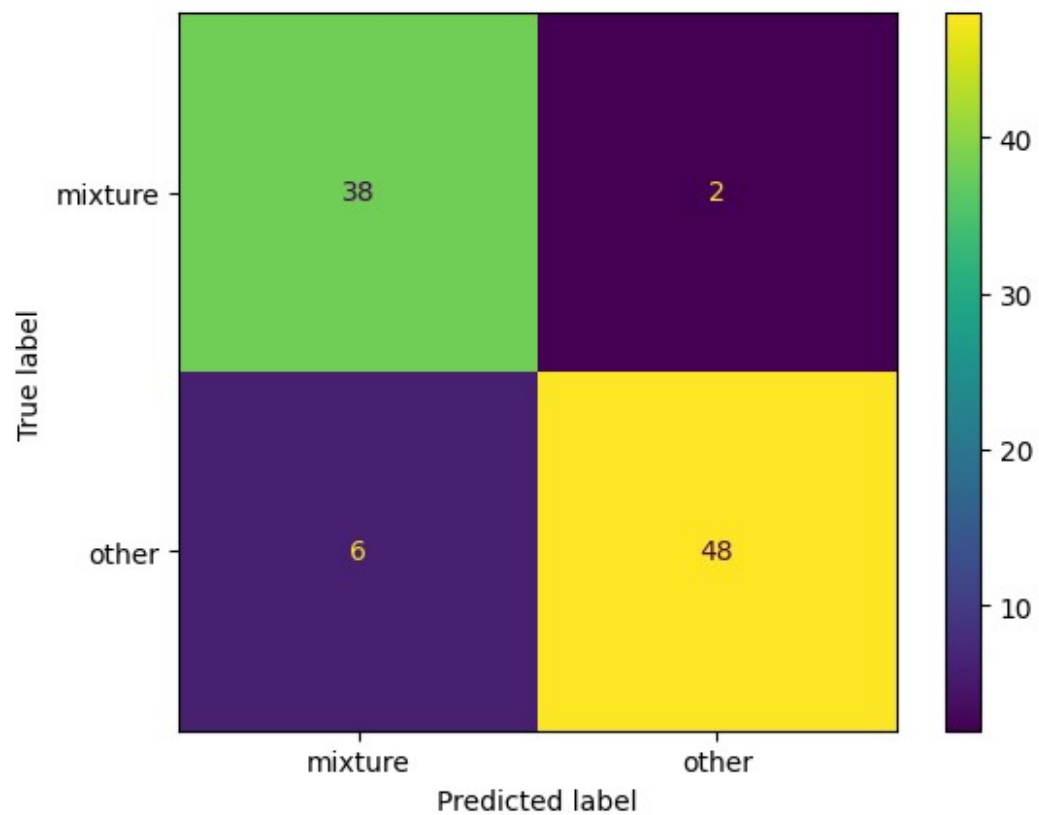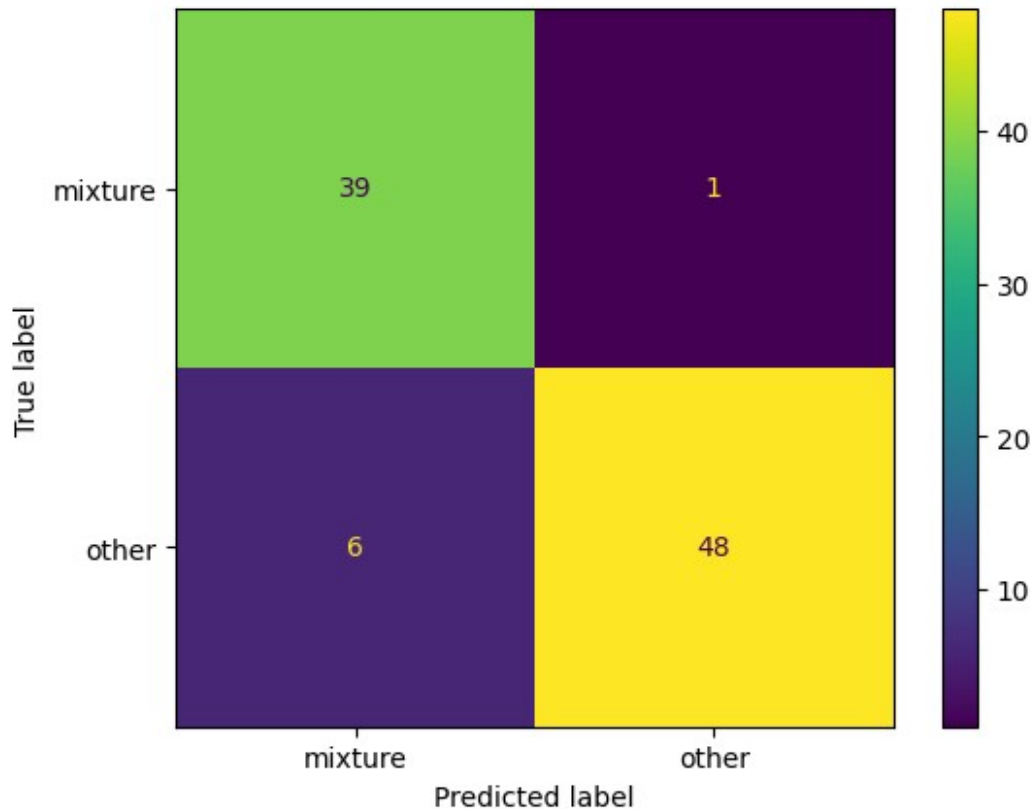
Etape 2 : Classification selon la colonne TEXT

On scinde notre jeu de données en jeu d'apprentissage et jeu de test

```
X=dftrain["text"]
print(X)

X_train,X_test,y_train,y_test=train_test_split(X,y,test_size =
0.2,random_state=8)
print("X_train is",X_train)
print("y_train is",y_train)
print("X_test is",X_test)
print("y_test is",y_test)
```

```
623     By Jim Francesconi As we pass the governance t...
717     Free schools are the 'modern engines of social...
2398    More than 11,000 criminal trials in England an...
676     On Tuesday, radio show host John Fredricks sta...
231     Duration 0:47 Villaraigosa talks high-speed ra...
                        ...
1217    RETRACTION, Feb. 21, 2019: This Op-Ed has been...
713     GETTY Leading economists are finally acknowled...
652     Iran Parliamentarians have moved a bill in ord...
889     Diabetes prescriptions are costing the NHS in ...
```

```
773      Jeb Bush Really Is His Own Man By Stephanie Sc...
Name: text, Length: 468, dtype: object
X_train is 816      Climate crisis and logging is leading to shift...
328      Madison, WI…Speaker Robin Vos R-Rochester and ...
2456     The far-left ThinkProgress reports that scient...
258      November 11, 2020 Victims of Voter Fraud: Init...
1959     Hundreds of Nigerian women die every day of co...
                          ...
1952     In response to Gen David Petraeus' denunciatio...
1518     Cash-strapped schools are spending a record £1...
1977     GETTY Leading economists are finally acknowled...
883      French researcher Dr. Didier Raoult has now tr...
168      The 2021 Act on Climate, legislation sponsored...
Name: text, Length: 374, dtype: object
y_train is 816        other
328      mixture
2456     mixture
258      mixture
1959     mixture
            ...
1952     mixture
1518     mixture
1977        other
883      mixture
168      mixture
Name: rating, Length: 374, dtype: object
X_test is 1390     As the giant hydra-armed machine that is the s...
424      There's not much to laugh about these days, bu...
2170     EXCLUSIVE: Third of cases won by terrorists, m...
1155     To read an annotated version of this article, ...
335      Florida Rep. Debbie Wasserman Schultz, the Dem...
                          ...
1209     Much of recent global warming has been fabrica...
599      After analysis of images shot by US-based spac...
1111     Get our daily coronavirus email newsletter wit...
2386     PUPILS aged just five have been accused of sex...
2385     L ondoners living in the borough of Greenwich ...
Name: text, Length: 94, dtype: object
y_test is 1390        other
424         other
2170        other
1155     mixture
335      mixture
            ...
1209     mixture
599         other
1111        other
2386        other
2385        other
Name: rating, Length: 94, dtype: object
```
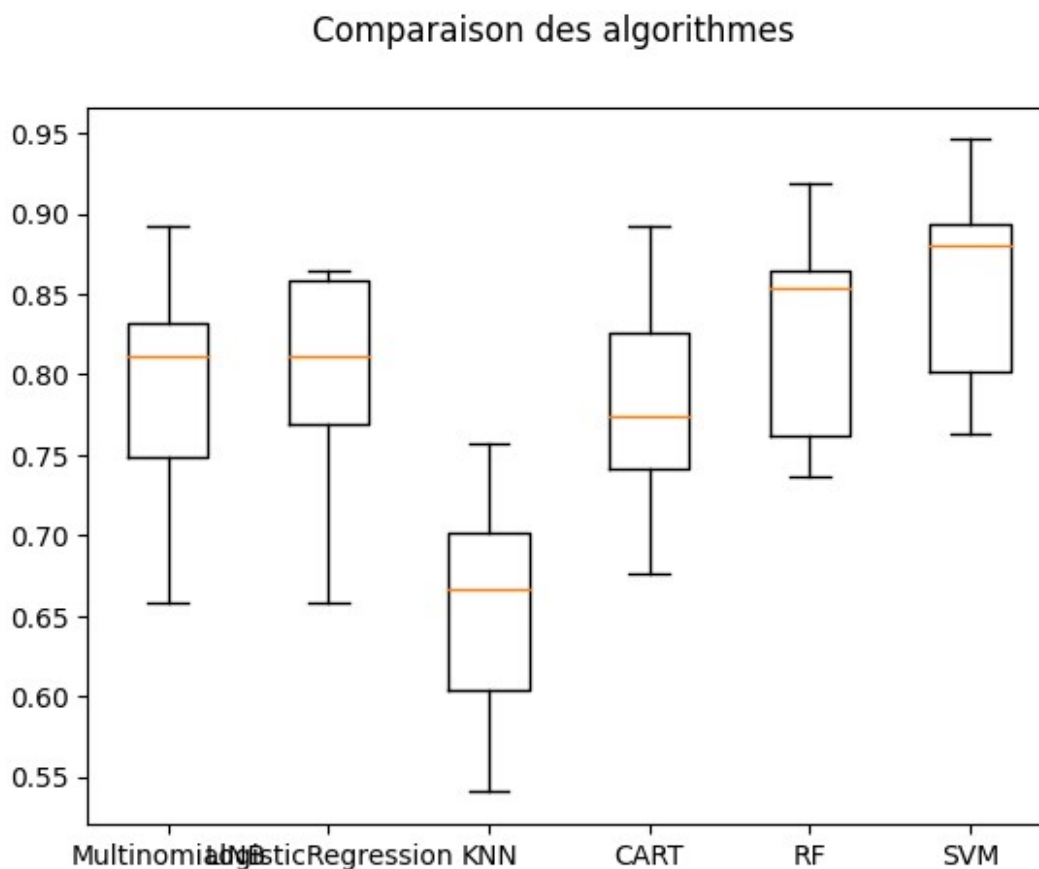
Affichage des boites à moustaches

```python
fig = plt.figure()
fig.suptitle('Comparaison des algorithmes')
ax = fig.add_subplot(111)
plt.boxplot(scores)
ax.set_xticklabels(names)
```

```
[Text(1, 0, 'MultinomialNB'),
 Text(2, 0, 'LogisticRegression'),
 Text(3, 0, 'KNN'),
 Text(4, 0, 'CART'),
 Text(5, 0, 'RF'),
 Text(6, 0, 'SVM')]
```


On a un pipeline pour chaque prétraitement différent, on essaye pas mal (miniscule, lemmatisation, miniscule + lemmatisation..) et on stocke le fit_transorm de nos X_train, X_test sur les pipelines dans des listes qui vont contenir tous les fit_transform des pipelines pour chaque classifieur, par la suite on parcourt ces listes là, on itère dessus, et chaque élement de la liste (train) va passer par le GridSearch et puis on predict sur son corresapondant dans liste (test).

```python
np.random.seed(42)  # Set the random seed for NumPy
# pipeline de l'utilisation de TfidfVectorizer avec differents pre-
```

```python
traitements
TFIDF_brut = Pipeline ([('cleaner', TextNormalizer()),
                        ('tfidf_vectorizer',
TfidfVectorizer(lowercase=False))])

TFIDF_lowcase = Pipeline([('cleaner',
TextNormalizer(removestopwords=False,lowercase=True,

getstemmer=False,removedigit=False)),
                        ('tfidf_vectorizer',
TfidfVectorizer(lowercase=False))])
TFIDF_lowStop = Pipeline([('cleaner',
TextNormalizer(removestopwords=True,lowercase=True,

getstemmer=False,removedigit=False)),
                        ('tfidf_vectorizer',
TfidfVectorizer(lowercase=False))])

TFIDF_lowStopstem = Pipeline([('cleaner',
TextNormalizer(removestopwords=True,lowercase=True,

getstemmer=True,removedigit=False)),
                        ('tfidf_vectorizer',
TfidfVectorizer(lowercase=False))])


# Liste de tous les modeles à tester
all_models = [
    ("TFIDF_lowcase", TFIDF_lowcase),
    ("TFIDF_lowStop", TFIDF_lowStop),
    ("TFIDF_lowStopstem",TFIDF_lowStopstem),
    ("TFIDF_brut", TFIDF_brut)
]

X_train_title_SVC = []
X_test_title_SVC = []



for name, pipeline in all_models :

X_train_title_SVC.append(pipeline.fit_transform(X_train).toarray())
    X_test_title_SVC.append(pipeline.transform(X_test).toarray())



models = {
    'SVC': SVC(random_state=42)
```

```python
}

params = {'SVC': [{'C': [0.001, 0.01, 0.1, 1,2]},
          {'gamma': [0.001, 0.01, 0.1,0.2,0.3]},
          {'kernel': ['linear', 'rbf']}]
}

for model_name, model in models.items():
    score='accuracy'
    X_train_title = eval('X_train_title_' + model_name)
    X_test_title = eval('X_test_title_' + model_name)
    for i in range (len(X_train_title)):
      grid_search = GridSearchCV(model, params[model_name], n_jobs=-1,
verbose=1,scoring=score)
      print("grid search fait")
      print("X_train",X_train_title[i].shape)
      print("y_train",y_train.shape)
      grid_search.fit(X_train_title[i],y_train)
      print ('meilleur score %0.3f'%(grid_search.best_score_),'\n')
      print ('meilleur estimateur',grid_search.best_estimator_,'\n')
      y_pred = grid_search.predict(X_test_title[i])
      MyshowAllScores(y_test,y_pred)

      print("Ensemble des meilleurs paramètres :")
      best_parameters = grid_search.best_estimator_.get_params()
      for param_dict in params[model_name]:
        for param_name, param_value in param_dict.items():
            print("\t%s: %r" % (param_name,
best_parameters[param_name]))
```

```
y_train (374,)
y_test (94,)
X_test (94,)
grid search fait
X_train (374, 15205)
y_train (374,)
Fitting 5 folds for each of 12 candidates, totalling 60 fits
meilleur score 0.845

meilleur estimateur SVC(C=1, random_state=42)

Accuracy : 0.915
Classification Report
              precision    recall   f1-score    support

     mixture    0.84783    0.97500    0.90698         40
       other    0.97917    0.87037    0.92157         54

    accuracy                          0.91489         94
```

```
    macro avg     0.91350    0.92269    0.91427          94
 weighted avg     0.92328    0.91489    0.91536          94


Ensemble des meilleurs paramètres :
     C: 1
     gamma: 'scale'
     kernel: 'rbf'
grid search fait
X_train (374, 15067)
y_train (374,)
Fitting 5 folds for each of 12 candidates, totalling 60 fits
meilleur score 0.850

meilleur estimateur SVC(C=1, random_state=42)

Accuracy : 0.926
Classification Report
               precision    recall   f1-score    support

      mixture    0.86667    0.97500    0.91765          40
        other    0.97959    0.88889    0.93204          54

     accuracy                          0.92553          94
    macro avg    0.92313    0.93194    0.92484          94
 weighted avg    0.93154    0.92553    0.92591          94


Ensemble des meilleurs paramètres :
     C: 1
     gamma: 'scale'
     kernel: 'rbf'
grid search fait
X_train (374, 10316)
y_train (374,)
Fitting 5 folds for each of 12 candidates, totalling 60 fits
meilleur score 0.845

meilleur estimateur SVC(C=1, random_state=42)

Accuracy : 0.926
Classification Report
               precision    recall   f1-score    support

      mixture    0.86667    0.97500    0.91765          40
        other    0.97959    0.88889    0.93204          54

     accuracy                          0.92553          94
    macro avg    0.92313    0.93194    0.92484          94
 weighted avg    0.93154    0.92553    0.92591          94
```

```
Ensemble des meilleurs paramètres :
      C: 1
      gamma: 'scale'
      kernel: 'rbf'
grid search fait
X_train (374, 17774)
y_train (374,)
Fitting 5 folds for each of 12 candidates, totalling 60 fits
meilleur score 0.842

meilleur estimateur SVC(C=1, random_state=42)

Accuracy : 0.915
Classification Report
              precision    recall  f1-score   support

     mixture    0.84783   0.97500   0.90698        40
       other    0.97917   0.87037   0.92157        54

    accuracy                        0.91489        94
   macro avg    0.91350   0.92269   0.91427        94
weighted avg    0.92328   0.91489   0.91536        94

Ensemble des meilleurs paramètres :
      C: 1
      gamma: 'scale'
      kernel: 'rbf'
```
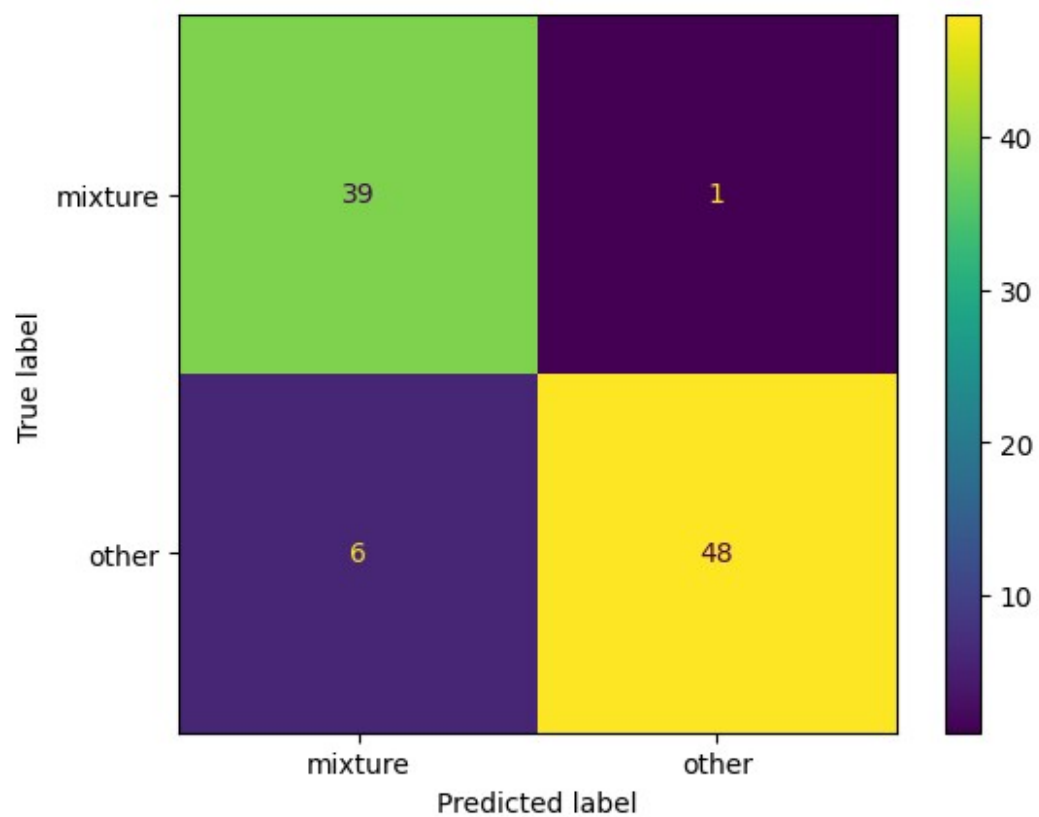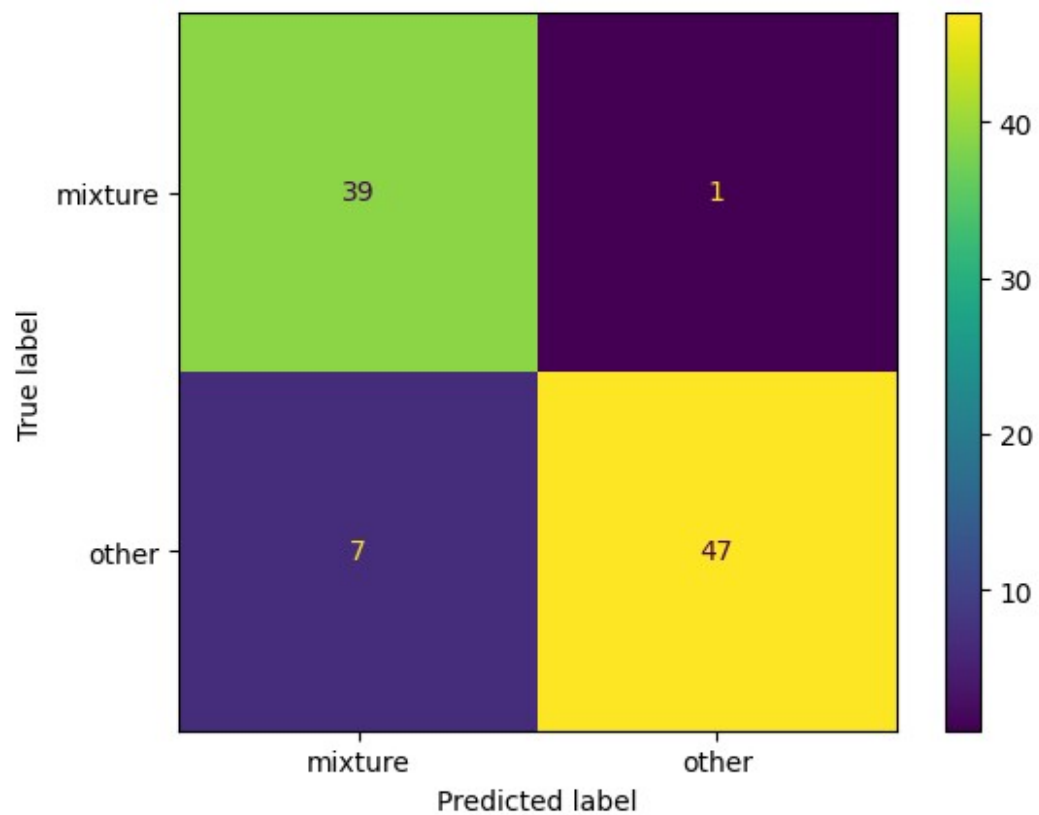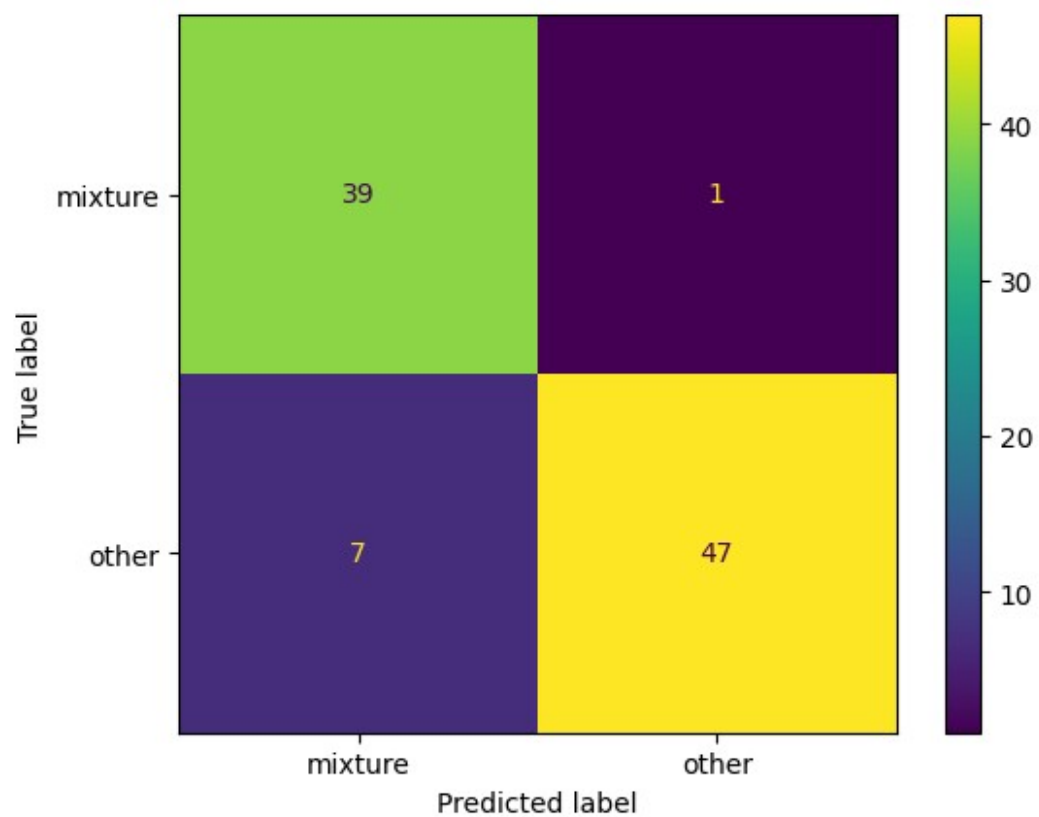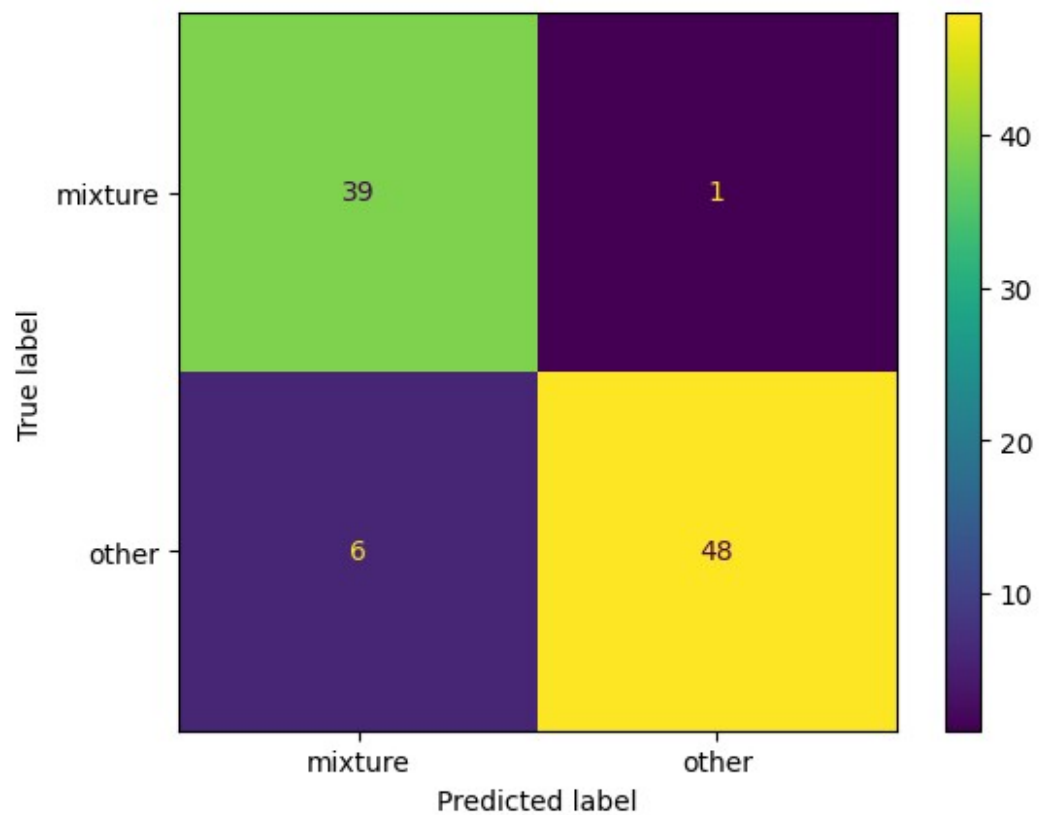
Classification selon la colonne TEXT_TITLE

On contactene la colonne titre et texte

```python
text_title = dftrain.apply(lambda x : '{}
{}'.format(x['text'],x['title']),axis=1)

dftrain['text_title'] = text_title
X_text_title = dftrain.iloc[0:,-1]
X_text_title.reset_index(drop = True, inplace = True)
```

On scinde notre jeu de données en jeu d'apprentissage et jeu de test

```python
X=dftrain["text_title"]
print(X)

X_train,X_test,y_train,y_test=train_test_split(X,y,test_size =
0.2,random_state=8)
print("X_train is",X_train)
print("y_train is",y_train)
print("X_test is",X_test)
print("y_test is",y_test)
```

```
623     By Jim Francesconi As we pass the governance t...
717     Free schools are the 'modern engines of social...
2398    More than 11,000 criminal trials in England an...
676     On Tuesday, radio show host John Fredricks sta...
231     Duration 0:47 Villaraigosa talks high-speed ra...
                             ...
1217    RETRACTION, Feb. 21, 2019: This Op-Ed has been...
713     GETTY Leading economists are finally acknowled...
652     Iran Parliamentarians have moved a bill in ord...
889     Diabetes prescriptions are costing the NHS in ...
773     Jeb Bush Really Is His Own Man By Stephanie Sc...
Name: text_title, Length: 468, dtype: object
X_train is 816       Climate crisis and logging is leading to shift...
328     Madison, WI…Speaker Robin Vos R-Rochester and ...
2456    The far-left ThinkProgress reports that scient...
258     November 11, 2020 Victims of Voter Fraud: Init...
1959    Hundreds of Nigerian women die every day of co...
                             ...
1952    In response to Gen David Petraeus' denunciatio...
1518    Cash-strapped schools are spending a record £1...
1977    GETTY Leading economists are finally acknowled...
883     French researcher Dr. Didier Raoult has now tr...
168     The 2021 Act on Climate, legislation sponsored...
Name: text_title, Length: 374, dtype: object
y_train is 816        other
328     mixture
2456    mixture
```

```
258       mixture
1959      mixture
           ...
1952      mixture
1518      mixture
1977       other
883       mixture
168       mixture
Name: rating, Length: 374, dtype: object
X_test is 1390    As the giant hydra-armed machine that is the s...
424       There's not much to laugh about these days, bu...
2170      EXCLUSIVE: Third of cases won by terrorists, m...
1155      To read an annotated version of this article, ...
335       Florida Rep. Debbie Wasserman Schultz, the Dem...
                              ...
1209      Much of recent global warming has been fabrica...
599       After analysis of images shot by US-based spac...
1111      Get our daily coronavirus email newsletter wit...
2386      PUPILS aged just five have been accused of sex...
2385      L ondoners living in the borough of Greenwich ...
Name: text_title, Length: 94, dtype: object
y_test is 1390        other
424        other
2170       other
1155      mixture
335       mixture
           ...
1209      mixture
599        other
1111       other
2386       other
2385       other
Name: rating, Length: 94, dtype: object
```

Ici, c'est une étape importante, on va tester différents classifieurs, pour chacun des classifieurs, on va appliquer le prétraitement + Vectorisation TfIdf, et on applique une cross_val_score avec un Kfold de 10 fois, par la suite on stocke dans une liste all_results la moyenne des accuracy + l'écart type et on la trie par ordre décroissant de moyenne d'accuracy et d'écart type. on remarque que les 2 meilleurs sont SVM et RF qu'on va séléctionner pour leur appliquer le GridSearch sur les paramètres des prétraitements + leurs hyperparamètres pour pouvoir choisir le meilleur.

```python
score = 'accuracy'
seed = 7
allresults = []
results = []
names = []


# Liste des modèles à tester
```

```python
models = [
    ('MultinomialNB', MultinomialNB()),
    ('LogisticRegression', LogisticRegression(random_state=42))
]


#models.append(('LR', LogisticRegression(solver='lbfgs')))
models.append(('KNN', KNeighborsClassifier()))
models.append(('CART', DecisionTreeClassifier(random_state=42)))
models.append(('RF', RandomForestClassifier(random_state=42)))
models.append(('SVM', SVC(random_state=42)))

# Création d'un pipeline pour chaque modèle
pipelines = []
for name,model in models:
    pipeline = Pipeline([
        ('normalize', TextNormalizer()),
        ('tfidf', TfidfVectorizer()),
        (name,model)
    ])
    pipelines.append((name,pipeline))

all_results=[]
scores=[]
for p in pipelines:
    print(p[1])
    # cross validation en 10 fois
    kfold = KFold(n_splits=10,random_state=seed,shuffle=True)

#     print ("Evaluation de ",p)
    start_time = time.time()
    # application de la classification
    cv_results = cross_val_score(p[1],X_train,y_train, cv=kfold,
scoring=score)
    scores.append(cv_results)
    names.append(p[0])
    all_results.append((p[0],cv_results.mean(),cv_results.std()))
    end_time = time.time()


print("all resultats", all_results)

all_results = sorted(all_results, key=lambda x: (-x[1], -x[2]))
print("all resultats", all_results)
    # affichage des résultats
#print ('\nLe meilleur resultat : ',max(results))

Pipeline(steps=[('normalize', TextNormalizer()), ('tfidf',
TfidfVectorizer()),
                ('MultinomialNB', MultinomialNB())])
```

```
Pipeline(steps=[('normalize', TextNormalizer()), ('tfidf',
TfidfVectorizer()),
                ('LogisticRegression',
LogisticRegression(random_state=42))])
Pipeline(steps=[('normalize', TextNormalizer()), ('tfidf',
TfidfVectorizer()),
                ('KNN', KNeighborsClassifier())])
Pipeline(steps=[('normalize', TextNormalizer()), ('tfidf',
TfidfVectorizer()),
                ('CART', DecisionTreeClassifier(random_state=42))])
Pipeline(steps=[('normalize', TextNormalizer()), ('tfidf',
TfidfVectorizer()),
                ('RF', RandomForestClassifier(random_state=42))])
Pipeline(steps=[('normalize', TextNormalizer()), ('tfidf',
TfidfVectorizer()),
                ('SVM', SVC(random_state=42))])
all resultats [('MultinomialNB', 0.7866287339971552,
0.0689717093735457), ('LogisticRegression', 0.8054054054054054,
0.07269404515568734), ('KNN', 0.6412517780938835,
0.04981222670527374), ('CART', 0.8131578947368421,
0.053777319993537714), ('RF', 0.8691322901849219,
0.04522151055040785), ('SVM', 0.8666429587482221,
0.06381925490389186)]
all resultats [('RF', 0.8691322901849219, 0.04522151055040785),
('SVM', 0.8666429587482221, 0.06381925490389186), ('CART',
0.8131578947368421, 0.053777319993537714), ('LogisticRegression',
0.8054054054054054, 0.07269404515568734), ('MultinomialNB',
0.7866287339971552, 0.0689717093735457), ('KNN', 0.6412517780938835,
0.04981222670527374)]
```

On affiche les accuracy de chaque classifieur, on remarque la médiane (en rouge) de chaque et l'écart type aussi.
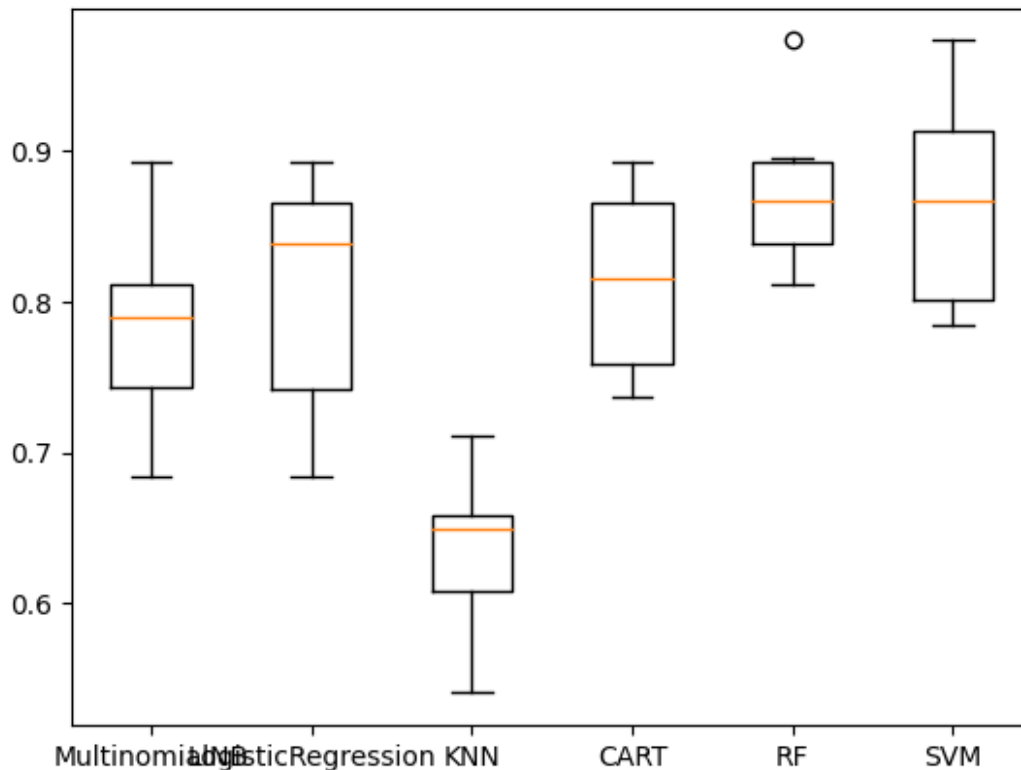
```python
import matplotlib.pyplot as plt
fig = plt.figure()
fig.suptitle('Comparaison des algorithmes')
ax = fig.add_subplot(111)
plt.boxplot(scores)
ax.set_xticklabels(names)
```

```
[Text(1, 0, 'MultinomialNB'),
 Text(2, 0, 'LogisticRegression'),
 Text(3, 0, 'KNN'),
 Text(4, 0, 'CART'),
 Text(5, 0, 'RF'),
 Text(6, 0, 'SVM')]
```

Comparaison des algorithmes

On a un pipeline pour chaque prétraitement différent, on essaye pas mal (miniscule, lemmatisation, miniscule + lemmatisation..) et on stocke le fit_transorm de nos X_train, X_test sur les pipelines dans des listes qui vont contenir tous les fit_transform des pipelines pour chaque classifieur, par la suite on parcourt ces listes là, on itère dessus, et chaque élement de la liste (train) va passer par le GridSearch et puis on predict sur son corresapondant dans liste (test).

```python
from sklearn.model_selection import GridSearchCV


from sklearn.datasets import fetch_20newsgroups
from sklearn.feature_extraction.text import CountVectorizer
from sklearn.feature_extraction.text import TfidfVectorizer
from sklearn.model_selection import train_test_split, GridSearchCV
from sklearn.pipeline import Pipeline
from sklearn.metrics import accuracy_score
from sklearn.naive_bayes import MultinomialNB
from sklearn.linear_model import LogisticRegression
from sklearn.svm import SVC
from sklearn.naive_bayes import MultinomialNB
from tabulate import tabulate
import numpy as np
np.random.seed(42)  # Set the random seed for NumPy
```

```python
print("y_train", y_train.shape)
print("y_test", y_test.shape)
#X_test = np.ravel(X_test)
print("X_test", X_test.shape)
# le plus simple est de faire un test sur differents pipelines.


# pipeline de l'utilisation de TfidfVectorizer avec differents pre-
traitements
TFIDF_brut = Pipeline ([('cleaner', TextNormalizer()),
                        ('tfidf_vectorizer',
TfidfVectorizer(lowercase=False))])

TFIDF_lowcase = Pipeline([('cleaner',
TextNormalizer(removestopwords=False,lowercase=True,

getstemmer=False,removedigit=False)),
                        ('tfidf_vectorizer',
TfidfVectorizer(lowercase=False))])
TFIDF_lowStop = Pipeline([('cleaner',
TextNormalizer(removestopwords=True,lowercase=True,

getstemmer=False,removedigit=False)),
                        ('tfidf_vectorizer',
TfidfVectorizer(lowercase=False))])

TFIDF_lowStopstem = Pipeline([('cleaner',
TextNormalizer(removestopwords=True,lowercase=True,

getstemmer=True,removedigit=False)),
                        ('tfidf_vectorizer',
TfidfVectorizer(lowercase=False))])


# Liste de tous les modeles à tester
all_models = [
    ("TFIDF_lowcase", TFIDF_lowcase),
    ("TFIDF_lowStop", TFIDF_lowStop),
    ("TFIDF_lowStopstem",TFIDF_lowStopstem),
    ("TFIDF_brut", TFIDF_brut)
]

X_train_title_SVC = []
X_test_title_SVC = []




for name, pipeline in all_models :
```

```python
X_train_title_SVC.append(pipeline.fit_transform(X_train).toarray())
    X_test_title_SVC.append(pipeline.transform(X_test).toarray())




models = {
    'SVC': SVC(random_state=42)
}


params = {'SVC': [{'C': [0.001, 0.01, 0.1, 1,2]},
            {'gamma': [0.001, 0.01, 0.1,0.2,0.3]},
            {'kernel': ['linear', 'rbf']}]
}

for model_name, model in models.items():
    score='accuracy'
    X_train_title = eval('X_train_title_' + model_name)
    X_test_title = eval('X_test_title_' + model_name)
    for i in range (len(X_train_title)):
        grid_search = GridSearchCV(model, params[model_name], n_jobs=-1,
verbose=1,scoring=score)
        print("grid search fait")
        print("X_train",X_train_title[i].shape)
        print("y_train",y_train.shape)
        grid_search.fit(X_train_title[i],y_train)
        print ('meilleur score %0.3f'%(grid_search.best_score_),'\n')
        print ('meilleur estimateur',grid_search.best_estimator_,'\n')
        y_pred = grid_search.predict(X_test_title[i])
        MyshowAllScores(y_test,y_pred)

        print("Ensemble des meilleurs paramètres :")
        best_parameters = grid_search.best_estimator_.get_params()
        for param_dict in params[model_name]:
          for param_name, param_value in param_dict.items():
                print("\t%s: %r" % (param_name,
best_parameters[param_name]))

y_train (374,)
y_test (94,)
X_test (94,)
grid search fait
X_train (374, 15575)
y_train (374,)
Fitting 5 folds for each of 12 candidates, totalling 60 fits
meilleur score 0.848

meilleur estimateur SVC(C=1, random_state=42)
```

```
Accuracy : 0.915
Classification Report
              precision    recall  f1-score   support

     mixture    0.84783   0.97500   0.90698        40
       other    0.97917   0.87037   0.92157        54

    accuracy                        0.91489        94
   macro avg    0.91350   0.92269   0.91427        94
weighted avg    0.92328   0.91489   0.91536        94
```

Ensemble des meilleurs paramètres :
```
     C: 1
     gamma: 'scale'
     kernel: 'rbf'
```
grid search fait
X_train (374, 15437)
y_train (374,)
Fitting 5 folds for each of 12 candidates, totalling 60 fits
meilleur score 0.853

meilleur estimateur SVC(C=2, random_state=42)

```
Accuracy : 0.926
Classification Report
              precision    recall  f1-score   support

     mixture    0.86667   0.97500   0.91765        40
       other    0.97959   0.88889   0.93204        54

    accuracy                        0.92553        94
   macro avg    0.92313   0.93194   0.92484        94
weighted avg    0.93154   0.92553   0.92591        94
```

Ensemble des meilleurs paramètres :
```
     C: 2
     gamma: 'scale'
     kernel: 'rbf'
```
grid search fait
X_train (374, 10551)
y_train (374,)
Fitting 5 folds for each of 12 candidates, totalling 60 fits
meilleur score 0.845

meilleur estimateur SVC(C=1, random_state=42)

```
Accuracy : 0.926
Classification Report
              precision    recall  f1-score   support
```

```
      mixture      0.86667      0.97500      0.91765             40
        other      0.97959      0.88889      0.93204             54

     accuracy                                0.92553             94
    macro avg      0.92313      0.93194      0.92484             94
 weighted avg      0.93154      0.92553      0.92591             94
```

Ensemble des meilleurs paramètres :
     C: 1
     gamma: 'scale'
     kernel: 'rbf'
grid search fait
X_train (374, 18379)
y_train (374,)
Fitting 5 folds for each of 12 candidates, totalling 60 fits
meilleur score 0.848

meilleur estimateur SVC(C=1, random_state=42)

Accuracy : 0.915
Classification Report

```
               precision      recall    f1-score    support

      mixture      0.84783      0.97500      0.90698             40
        other      0.97917      0.87037      0.92157             54

     accuracy                                0.91489             94
    macro avg      0.91350      0.92269      0.91427             94
 weighted avg      0.92328      0.91489      0.91536             94
```
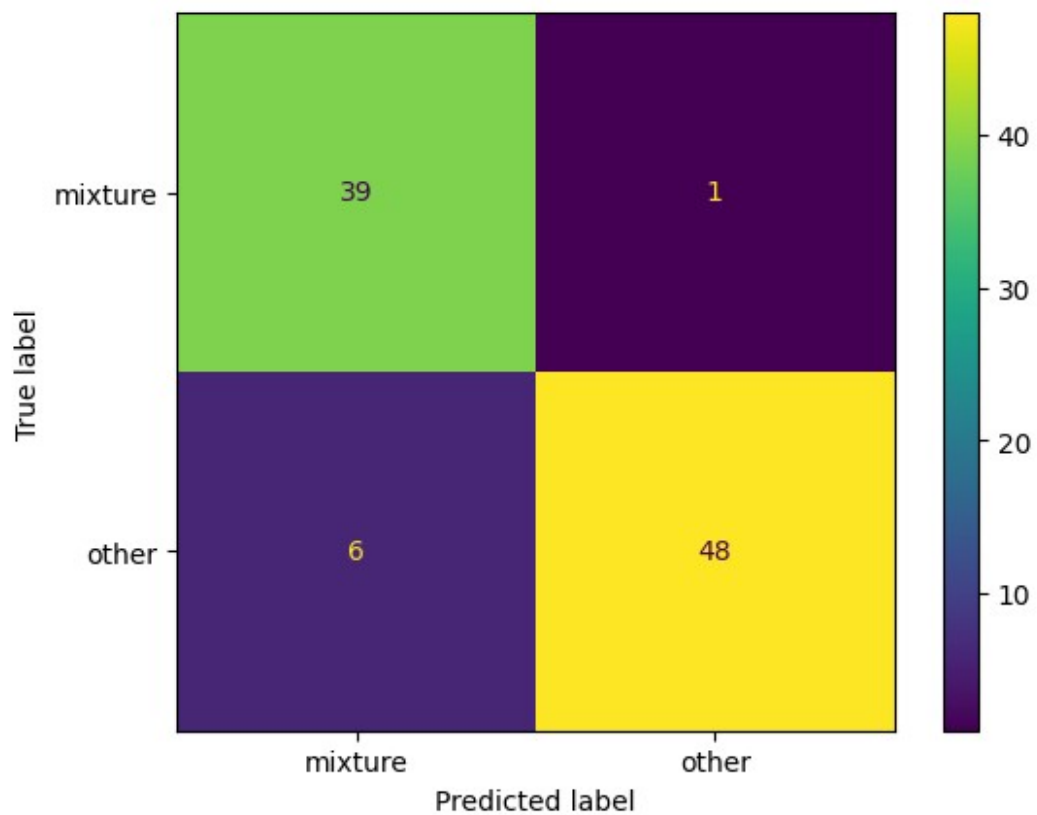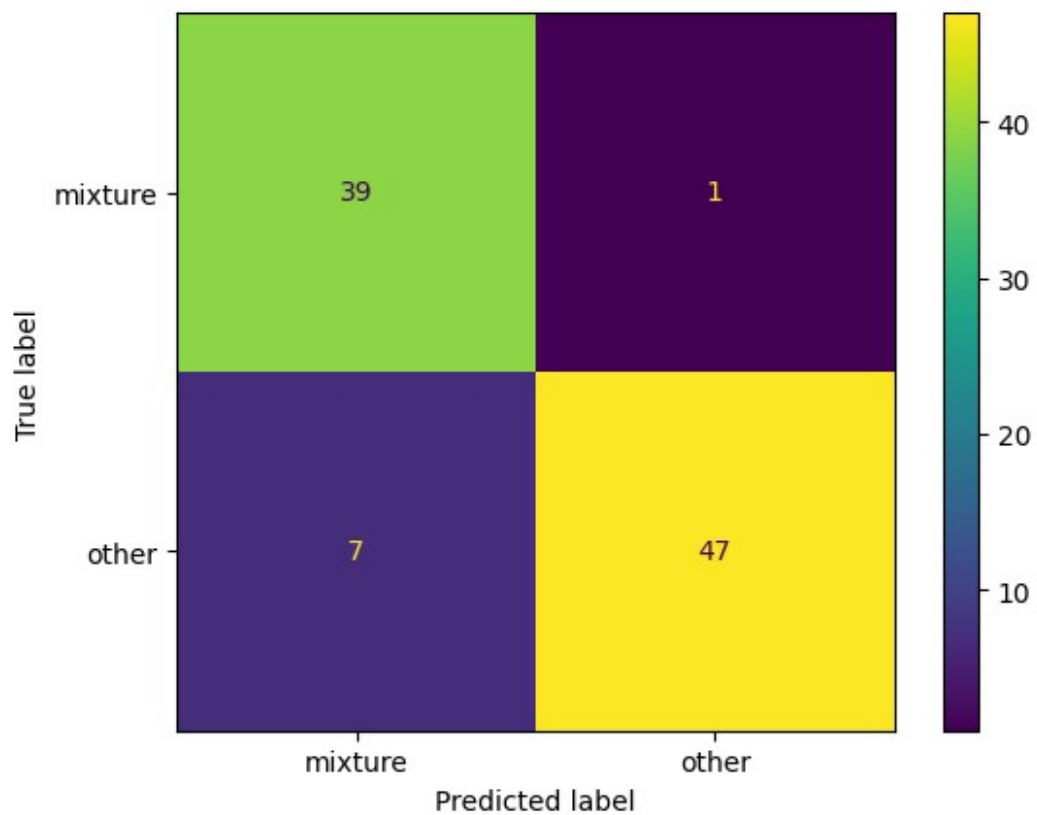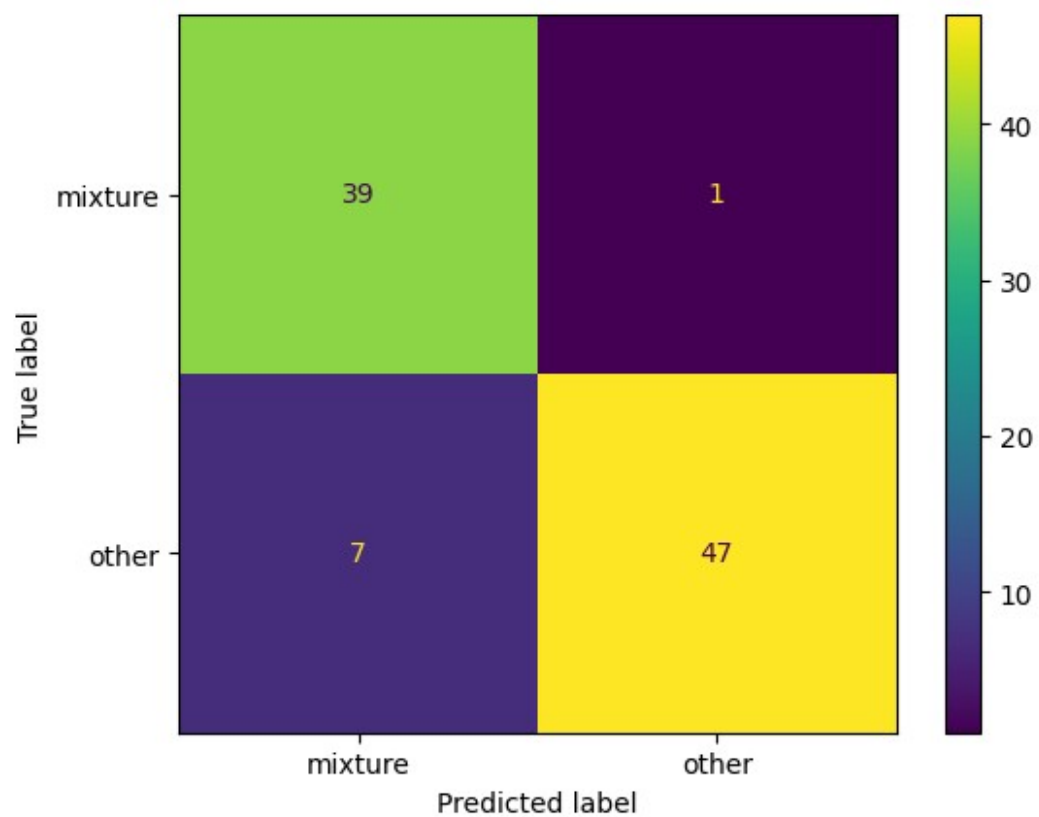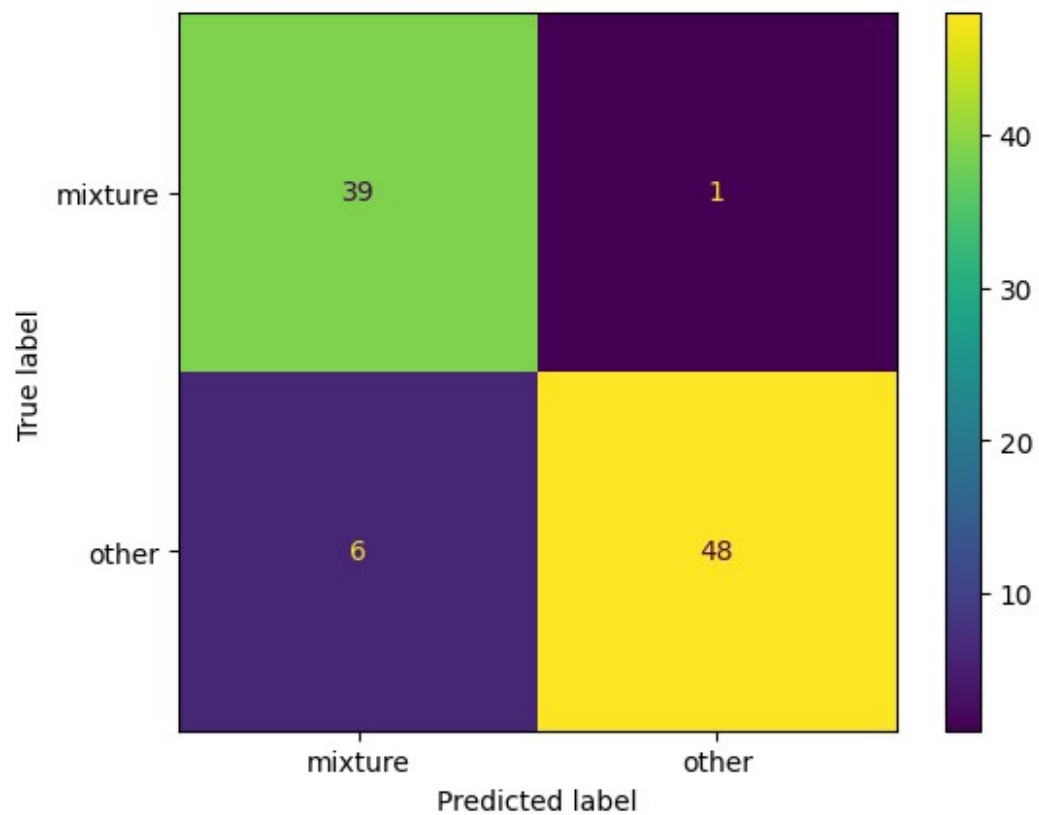
Ensemble des meilleurs paramètres :
     C: 1
     gamma: 'scale'
     kernel: 'rbf'

Classification de FALSE/TRUE avec les meilleurs prétraitements et meilleurs paramètres

```python
dftrain = dftrainbase.loc[dftrainbase['rating'].isin(['TRUE',
'FALSE'])]
y=dftrain.iloc[0:,-2]
print(y)
print("les valeurs de TRUE et FALSE maintenant sont
" ,y.value_counts())
```

```
0       FALSE
3       FALSE
4       FALSE
6       FALSE
7       FALSE
        ...
2523     TRUE
2524     TRUE
2525    FALSE
2526     TRUE
2527     TRUE
Name: rating, Length: 1578, dtype: object
les valeurs de TRUE et FALSE maintenant sont  FALSE    1156
TRUE      422
Name: rating, dtype: int64
```

```python
# Compter le nombre d'observations dans chaque catégorie
df_true = dftrain[dftrain['rating']=="TRUE"]
df_false = dftrain[dftrain['rating']=="FALSE"]

# Sous-échantillonner la classe majoritaire (FALSE) pour obtenir un
nombre égal d'échantillons pour chaque classe
df_false_subsampled = df_false.sample(n=len(df_true), random_state=42)

# Concaténer les deux dataframes
dftrain = pd.concat([df_false_subsampled, df_true])

# Mélanger aléatoirement les données
dftrain = dftrain.sample(frac=1, random_state=42)
```

```python
X_text=dftrain["text"]
X_title=dftrain["title"]
print("La taille de X_text",X_text.shape)
y=dftrain.iloc[0:,-2]
print(y.shape)
print("\n")
print("la taille de y est " ,y.shape)
print("\n")
```

```python
print("les valeurs de TRUE et FALSE maintenant sont
" ,y.value_counts())
```

La taille de X_text (844,)
(844,)

la taille de y est (844,)

les valeurs de TRUE et FALSE maintenant sont TRUE   422
FALSE   422
Name: rating, dtype: int64

```python
dftrain['text'] = dftrain['text'].apply(add_entity_name)
```


```python
X=dftrain.iloc[0:, 1:3]
print(X)
```

```python
X_train,X_test,y_train,y_test=train_test_split(X,y,test_size =
0.2,random_state=8)
print("X_train is",X_train)
print("y_train is",y_train)
print("X_test is",X_test)
print("y_test is",y_test)
y_test_1 = y_test
```

```
                                                  text  \
615   It 's been a long time coming , but finally we...
1303  Constitutional Attorney Matthew (PERSON) DePer...
1232  The (GPE) United (GPE) States (GPE) is witness...
2022  After three (DATE) decades (DATE) on the bench...
287   Based on actual results and accounting for sta...
...                                                 ...
1006  5 (MONEY) Million (MONEY) Muslim (MONEY) Child...
1543  The bombshell claim comes from over (TIME) 20 ...
853   BILL (ORG) GATES (ORG) EXPLAINS (ORG) THAT (OR...
296   Let our journalists help you make sense of the...
1325  Though the whole world relies on RT (ORG) - (O...

                                                 title
615    JK Rowling Confirms Stance Against Transgender...
1303   MI Sec of State Official Caught On Video Telli...
1232   What science can tell us about the links betwe...
2022         Sarah Parker leaves legacy on Supreme Court
287    Current Actual Election Result Update: Preside...
...                                                  ...
1006   Re: Meeting the need for isolation space for h...
1543   Breaking: Breonna Taylor 's boyfriend says SHE ...
```

```
853    A quote from Politifact: Gates never said that...
296    Before This Election, Newt Gingrich Believed t...
1325   COVID19 PCR Tests are Scientifically Meaningless

[844 rows x 2 columns]
X_train is                                                    text  \
1867  ( CNN (ORG) ) From the moment a child is born ...
1540  During the (ORG) Republican (ORG) National (OR...
2495  A large research synthesis , published in one ...
1197  Global warming could be far worse than predict...
788   Nearly (CARDINAL) 40,000 (CARDINAL) Wisconsini...
...                                                    ...
2224  TIJUANA , Mexico (GPE) — It 's the image from ...
2424  The scale of Antarctica (LOC) is startling . M...
2038  An 18 (DATE) - (DATE) year (DATE) - (DATE) old...
1295  It is absolutely right that across this House ...
1890  Justine (PERSON) Greening (PERSON) resigned fr...

                                                    title
1867   Trump's Tariffs Look Like a Self-Inflicted Wound
1540       Biden Says the Second Amendment is 'Obsolete'
2495  Scientists have just detected a major change t...
1197  Global warming could be far worse than predict...
788   "Anyone raising concern about the safety of Co...
...                                                    ...
2224  A discussion of 'smokers' black lungs' started...
2424                 Miles of Ice Collapsing Into the Sea
2038                                                   NaN
1295                              Wednesday 25 April 2018
1890  CLARKSBURG — When Jared Henry and Maddie Manse...

[675 rows x 2 columns]
y_train is 1867     TRUE
1540     FALSE
2495      TRUE
1197      TRUE
788       TRUE
         ...
2224      TRUE
2424      TRUE
2038      TRUE
1295      TRUE
1890      TRUE
Name: rating, Length: 675, dtype: object
X_test is                                                    text  \
360   Then - Secretary of State (ORG) Hillary (PERSO...
1216  Story highlights Global (LOC) sea (LOC) level ...
2466  Australia (GPE) 's Great (WORK_OF_ART) Barrier...
1757  News| [ email protected ] " If you wo n't lead...
1387  For once , it 's not President Trump (PERSON) ...
```

```
...                                                    ...
2435   The oldest and thickest sea ice in the Arctic ...
523    A coalition of civil society groups has descri...
2127   A coalition of civil society groups has descri...
247    With merchants in Democrat (NORP) - run cities...
615    It 's been a long time coming , but finally we...


                                                      title
360             Coverup on Benghazi does make a difference
1216   Observations show sea levels rising, and clima...
2466   Australia's Great Barrier Reef has worst coral...
1757   IT'S OFFICIAL: Brexit Britain WILL thrive out ...
1387   Marco Rubio Just Publicly Shamed His Own Const...
...                                                    ...
2435   Arctic's strongest sea ice breaks up for first...
523    Tesco food waste rose to equivalent of 119m me...
2127                                                    NaN
247    Antifa gearing up for false flag violence disg...
615    JK Rowling Confirms Stance Against Transgender...

[169 rows x 2 columns]
y_test is 360       TRUE
1216      TRUE
2466      TRUE
1757     FALSE
1387     FALSE
          ...
2435      TRUE
523       TRUE
2127      TRUE
247      FALSE
615       TRUE
Name: rating, Length: 169, dtype: object
```

```python
# création du pipeline en ajoutant le classifier
pipe = Pipeline([('cleaner',
TextNormalizer(removestopwords=True,lowercase=True,

getstemmer=True,removedigit=False)),
                ('tfidf_vectorizer',
TfidfVectorizer(lowercase=False)),
                ("SVM", SVC(C=2, gamma='scale', kernel='rbf',
random_state=42))])
pipetexte=pipe.fit(X_train["text"],y_train)
print("à présent",y_train)

#save pipe
print("pipeline créé")
```

```
à présent 1867      TRUE
1540      FALSE
2495       TRUE
1197       TRUE
788        TRUE

          ...
2224       TRUE
2424       TRUE
2038       TRUE
1295       TRUE
1890       TRUE
Name: rating, Length: 675, dtype: object
pipeline créé
```

Résultats de la classification selon True et False

```
X_test_text=X_test['text']
print(X_test_text)
y_pred_1=pipetexte.predict(X_test_text)
MyshowAllScores(y_test,y_pred_1)
```

```
360      Then - Secretary of State (ORG) Hillary (PERSO...
1216     Story highlights Global (LOC) sea (LOC) level ...
2466     Australia (GPE) 's Great (WORK_OF_ART) Barrier...
1757     News| [ email protected ] " If you wo n't lead...
1387     For once , it 's not President Trump (PERSON) ...
                          ...
2435     The oldest and thickest sea ice in the Arctic ...
523      A coalition of civil society groups has descri...
2127     A coalition of civil society groups has descri...
247      With merchants in Democrat (NORP) - run cities...
615       It 's been a long time coming , but finally we...
Name: text, Length: 169, dtype: object
Accuracy : 0.941
Classification Report
              precision    recall  f1-score   support

       FALSE    0.92941   0.95181   0.94048        83
        TRUE    0.95238   0.93023   0.94118        86

    accuracy                        0.94083       169
   macro avg    0.94090   0.94102   0.94083       169
weighted avg    0.94110   0.94083   0.94083       169
```
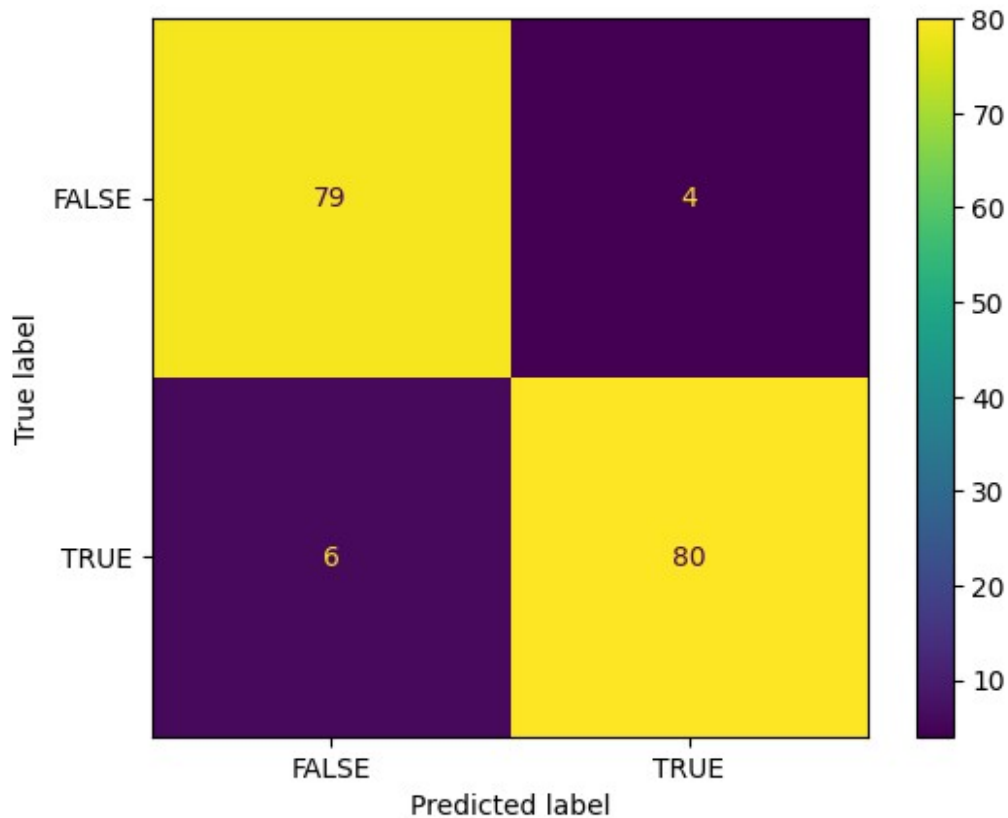
Classification de OTHER/MIXTURE avec les meilleurs prétraitements et meilleurs paramètres

```
dftrain = dftrainbase.loc[dftrainbase['rating'].isin(['other',
'mixture'])]
y=dftrain.iloc[0:,-2]
print(y)
print("les valeurs de MIXTURE et OTHER maintenant sont
" ,y.value_counts())
```

```
1        mixture
2        mixture
5          other
10       mixture
13       mixture
          ...
2517     mixture
2518     mixture
2519     mixture
2520     mixture
2522       other
Name: rating, Length: 950, dtype: object
les valeurs de MIXTURE et OTHER maintenant sont  mixture    716
other        234
Name: rating, dtype: int64
```

```python
# Compter le nombre d'observations dans chaque catégorie
mixture_count = dftrain['rating'].value_counts()['mixture']
other_count = dftrain['rating'].value_counts()['other']

# Trouver le nombre minimum d'observations parmi les catégories
min_count = min( mixture_count, other_count)

# Sous-échantillonner les catégories pour équilibrer les quantités
mixture_sampled = dftrain[dftrain['rating'] ==
'mixture'].sample(min_count, random_state=42)
other_sampled = dftrain[dftrain['rating'] ==
'other'].sample(min_count, random_state=42)


# Concaténer les échantillons pour obtenir un nouveau dataframe
équilibré
dftrain = pd.concat([mixture_sampled,other_sampled])

# Mélanger aléatoirement les données
dftrain = dftrain.sample(frac=1, random_state=42)
X_text=dftrain["text"]
X_title=dftrain["title"]
y=dftrain.iloc[0:,-2]
print("\n")
print("la taille de y_train est " ,y.shape)
print("\n")
print("les valeurs de OTHER et MIXTURE maintenant sont
" ,y.value_counts())
```

```
la taille de y_train est  (468,)


les valeurs de OTHER et MIXTURE maintenant sont  mixture     234
other       234
Name: rating, dtype: int64
```

```python
dftrain['title'] = dftrain['title'].apply(add_entity_name)


X=dftrain["title"]
print(X)

X_train,X_test,y_train,y_test=train_test_split(X,y,test_size =
0.2,random_state=8)
print("X_train is",X_train)
print("y_train is",y_train)
print("X_test is",X_test)
```

```
print("y_test is",y_test)
y_test_2 = y_test

623     DUP man dating Poots ' daughter claims his ' t...
717     Kremlin (ORG) calls new U.S. (GPE) sanctions i...
2398    Thousands (CARDINAL) of trials end over no - s...
676     Warren (PERSON) Statement (PERSON) on Boeing (...
231     Antonio (PERSON) Villaraigosa (PERSON) rips Ga...
                              ...
1217    [ RETRACTED ] Media Hysteria : Climate Change ...
713     With Stickers , a Petition and Even a Middle N...
652     Opportunities Foreclosed Equity Denied New (GP...
889     Renationalising railways supported by almost (...
773     Putin (PERSON) oversees hypersonic weapon test...
Name: title, Length: 468, dtype: object
X_train is 816     Zakharova (ORG) Says New (ORG) Salisbury (ORG)...
328     Rep. Vos (PERSON) , Sen. Fitzgerald (PERSON) :...
2456    Nolte (PERSON) : Scientists Prove Man - Made G...
258     Victims of Voter Fraud : Initial (WORK_OF_ART)...
1959    18 (DATE) - (DATE) Year (DATE) - (DATE) Old (D...
                              ...
1952    Global Warming Is Driving Polar Bears Toward E...
1518    Schools (ORG) spending £ 1.3bn (MONEY) on supp...
1977    With Stickers , a Petition and Even a Middle N...
883     Florida (GPE) Police Issue First Order to Disa...
168     State of Rhode (ORG) Island (ORG) General (ORG...
Name: title, Length: 374, dtype: object
y_train is 816         other
328     mixture
2456    mixture
258     mixture
1959    mixture
          ...
1952    mixture
1518    mixture
1977      other
883     mixture
168     mixture
Name: rating, Length: 374, dtype: object
X_test is 1390    Biden (ORG) Accepts (ORG) Campaign (ORG) Donat...
424                      Smoke (NORP) fags , save lives
2170    European (ORG) Court (ORG) of (ORG) Killers (O...
1155    When Will The (WORK_OF_ART) Planet (WORK_OF_AR...
335     Wasserman (PERSON) Schultz (PERSON) considerin...
                              ...
1209    DELINGPOLE : ' Nearly All ' Recent Global Warm...
599     Man claims there is a ' GHOST ' of a little gi...
1111    Kids do n't play a ' significant role ' in spr...
2386    Pervs (ORG) ' aged five (CARDINAL) School (ORG...
2385    London (GPE) 's most miserable boroughs reveal...
```

```
Name: title, Length: 94, dtype: object
y_test is 1390        other
424        other
2170        other
1155     mixture
335     mixture
           ...
1209     mixture
599        other
1111        other
2386        other
2385        other
Name: rating, Length: 94, dtype: object
```

```python
# création du pipeline en ajoutant le classifier
pipe =  Pipeline ([('cleaner', TextNormalizer()),
                   ('tfidf_vectorizer',
TfidfVectorizer(lowercase=False)),
                   ("SVM", SVC(C=2, gamma='scale', kernel='rbf',
random_state=42))])
pipetexte=pipe.fit(X_train,y_train)
print("à présent",y_train)

#save pipe
print("pipeline créé")
```

```
à présent 816        other
328     mixture
2456     mixture
258     mixture
1959     mixture
           ...
1952     mixture
1518     mixture
1977        other
883     mixture
168     mixture
Name: rating, Length: 374, dtype: object
pipeline créé
```

```python
X_test_text=X_test
print(X_test_text)
y_pred_2=pipetexte.predict(X_test_text)
MyshowAllScores(y_test,y_pred_2)
```

```
1390     Biden (ORG) Accepts (ORG) Campaign (ORG) Donat...
424                        Smoke (NORP) fags , save lives
2170     European (ORG) Court (ORG) of (ORG) Killers (O...
1155     When Will The (WORK_OF_ART) Planet (WORK_OF_AR...
335     Wasserman (PERSON) Schultz (PERSON) considerin...
                            ...
```

```
1209     DELINGPOLE : ' Nearly All ' Recent Global Warm...
599      Man claims there is a ' GHOST ' of a little gi...
1111     Kids do n't play a ' significant role ' in spr...
2386     Pervs (ORG) ' aged five (CARDINAL) School (ORG...
2385     London (GPE) 's most miserable boroughs reveal...
Name: title, Length: 94, dtype: object
Accuracy : 0.926
Classification Report
              precision    recall  f1-score   support

     mixture    0.86667   0.97500   0.91765        40
       other    0.97959   0.88889   0.93204        54

    accuracy                        0.92553        94
   macro avg    0.92313   0.93194   0.92484        94
weighted avg    0.93154   0.92553   0.92591        94
```
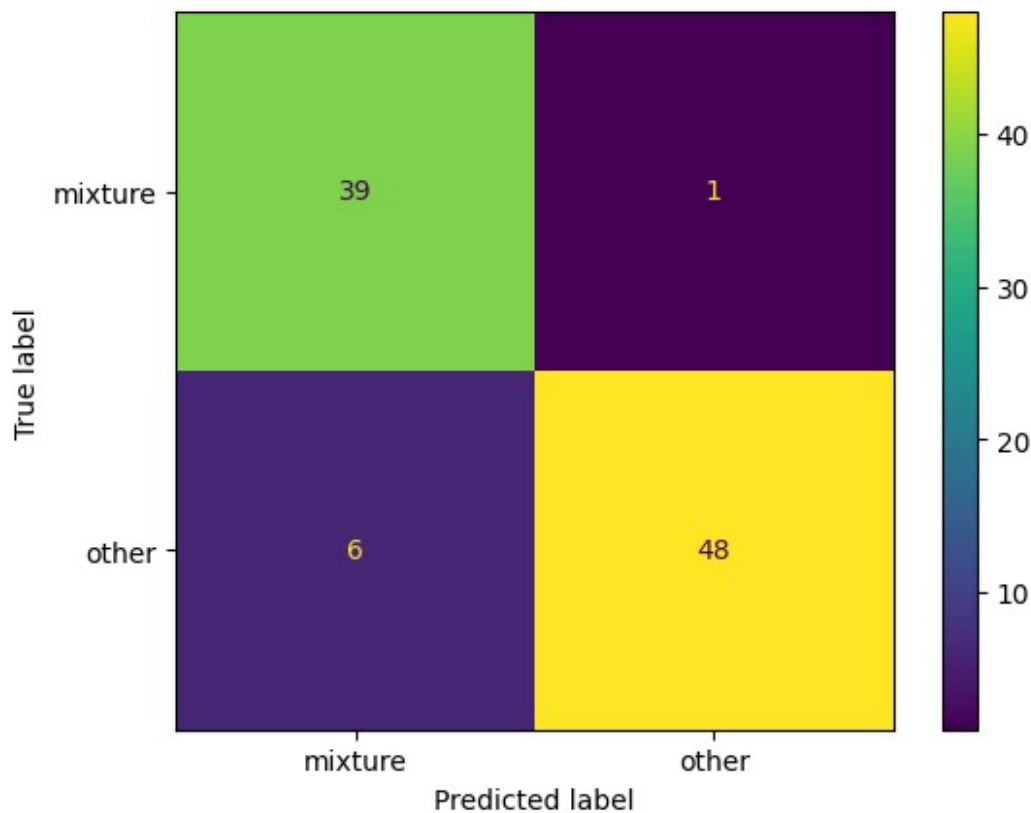

##Concaténons les résultats des deux classifications

```
# Conversion de tableaux NumPy en objets Series Pandas
y_pred_1 = pd.Series(y_pred_1)
y_pred_2 = pd.Series(y_pred_2)
#concat
y_pred_final = pd.concat([y_pred_1, y_pred_2], ignore_index=True)
```

```python
y_test_final = pd.concat([y_test_1, y_test_2], ignore_index=True)
#dernier test
MyshowAllScores(y_test_final,y_pred_final)
```

```
Accuracy : 0.935
Classification Report
              precision    recall  f1-score   support

       FALSE    0.92941   0.95181   0.94048        83
        TRUE    0.95238   0.93023   0.94118        86
     mixture    0.86667   0.97500   0.91765        40
       other    0.97959   0.88889   0.93204        54

    accuracy                        0.93536       263
   macro avg    0.93201   0.93648   0.93283       263
weighted avg    0.93768   0.93536   0.93550       263
```