```
Prédiction de l'obtention d'un permis avec un réseau de neurones artificiel
      1. Nettoyage du dataset
In [89]: # Importation des librairies
      import pandas as pd
      import numpy as np
      import warnings
      warnings.filterwarnings('ignore')
      # Importtion du dataset
      df_permit = pd.read_csv('dataset/Building_Permits.csv')
      # Suppression des catégories inutiles
      df_permit = df_permit.drop(['Permit Number'],axis=1)
      df_permit = df_permit.drop(['Block'],axis=1)
      df_permit = df_permit.drop(['Lot'],axis=1)
      df_permit = df_permit.drop(['Street Number'],axis=1)
      df_permit = df_permit.drop(['Street Number Suffix'],axis=1)
      df_permit = df_permit.drop(['Unit'],axis=1)
      df_permit = df_permit.drop(['Unit Suffix'],axis=1)
      df_permit = df_permit.drop(['Description'],axis=1)
      df_permit = df_permit.drop(['Permit Expiration Date'],axis=1)
      df_permit = df_permit.drop(['Estimated Cost'],axis=1)
      df permit = df_permit.drop(['Existing Use'],axis=1)
      df_permit = df_permit.drop(['Existing Units'],axis=1)
      df_permit = df_permit.drop(['Plansets'],axis=1)
      df_permit = df_permit.drop(['Location'],axis=1)
      df permit = df_permit.drop(['Record ID'],axis=1)
      df_permit = df_permit.drop(['Filed Date'],axis=1)
      #Remplacement des données manquantes avec des zéros (voir paramètre du fillna)
      df_permit['Number of Existing Stories'] = df_permit['Number of Existing Stories'].fillna(0)
      df_permit['Number of Proposed Stories'] = df_permit['Number of Proposed Stories'].fillna(0)
      df_permit['Structural Notification'] = df_permit['Structural Notification'].fillna(0)
      df_permit['Voluntary Soft-Story Retrofit'] = df_permit['Voluntary Soft-Story Retrofit'].fillna(0)
      df_permit['Fire Only Permit'] = df_permit['Fire Only Permit'].fillna(0)
      df_permit['TIDF Compliance'] = df_permit['TIDF Compliance'].fillna(0)
      df_permit['Site Permit'] = df_permit['Site Permit'].fillna(0)
      df_permit['Street Suffix'] = df_permit['Street Suffix'].fillna(0)
      df_permit['Existing Construction Type'] = df_permit['Existing Construction Type'].fillna(0)
      df_permit['Proposed Construction Type'] = df_permit['Proposed Construction Type'].fillna(0)
      df_permit['Existing Construction Type Description'] = df_permit['Existing Construction Type Description'].fillna(0)
      df_permit['Proposed Construction Type Description'] = df_permit['Proposed Construction Type Description'].fillna(0)
      df_permit['Proposed Use'] = df_permit['Proposed Use'].fillna(0)
      #Remplacement des données manquantes avec la médiane (voir paramètre du fillna)
      df_permit['Revised Cost'] = df_permit['Revised Cost'].fillna(np.nanmedian(df_permit['Revised Cost']))
      df_permit['Proposed Units'] = df_permit['Proposed Units'].fillna(np.nanmedian(df_permit['Proposed Units']))
      #Suppression des données manquantes
      df_permit = df_permit.dropna(subset=['First Construction Document Date'])
      df_permit = df_permit.dropna(subset=['Neighborhoods - Analysis Boundaries'])
      df_permit = df_permit.dropna(subset=['Zipcode'])
      df_permit = df_permit.dropna(subset=['Issued Date'])
      df_permit.head()
Out[89]:
                                                                    First
                                            Current
                      Permit
                                                                                           Exist
               Permit
                                                   Issued | Completed | Construction
                            Street | Street |
                                     Current
        Permit
                                                                          Proposed
                                                                                     TIDF
                                                                                        Construct
                Type
                     Creation
                                            Status
                                                                            Units Compliance
                            Name | Suffix
                                     Status
                                                                 Document
         Type
                                                     Date
                                                            Date
                                             Date
             Definition
                       Date
                                                                                             Ty
                                                                    Date
             sign - erect | 05/06/2015 | Ellis
                                St
                                          12/21/2017 | 11/09/2015 | NaN
                                                                11/09/2015
                                                                         2.0
                                                                                        3.0
                                    expired
             sign - erect | 04/19/2016 | Geary | St
                                          | 08/03/2017 | 08/03/2017 | NaN
                                                               08/03/2017
                                                                         2.0
                                                                                        3.0
                                    issued
                    complete | 07/24/2017 | 07/18/2017 | 07/24/2017 | 07/18/2017
                                                                        .. 1.0
                                                                                        5.0
             alterations
             permit
      4 6
             demolitions 11/28/2016 Market St
                                          12/01/2017 | 12/01/2017 | NaN
                                                               11/20/2017
                                                                         2.0
                                                                                        3.0
                                    issued
             otc
                   06/14/2017 | Indiana | St
                                          07/06/2017 07/06/2017 NaN
                                                                          326.0
                                                                                        1.0
      5 8
             alterations
                                                               07/06/2017
                                    lissued
             permit
      5 rows × 27 columns
In [90]: # Modification des colonnes 'date' :
      # - Création permit date devient 0
      # - Les autres dates sont remplacées par le nombre de jour écoulé depuis la création du permis
      df_date = df_permit[['Permit Creation Date','Current Status Date','Issued Date','Completed Date']]
      df_date['Permit Creation Date'] = pd.to_datetime(df_date['Permit Creation Date'])
      df_date['Current Status Date'] = pd.to_datetime(df_date['Current Status Date'])
      df_date['Issued Date'] = pd.to_datetime(df_date['Issued Date'])
      df_date['Completed Date'] = df_date['Completed Date'].fillna("12/12/2050")
      df_date['Completed Date'] = pd.to_datetime(df_date['Completed Date'])
      df_date2=df_date
      df_date2['Issued Date'] = (df_date['Issued Date'] - df_date['Permit Creation Date']).dt.days
      df_date2['Current Status Date'] = (df_date2['Current Status Date'] - df_date2['Permit Creation Date']).dt.days
      df_date2['Completed Date'] = (df_date2['Completed Date'] - df_date2['Permit Creation Date']).dt.days
      df_date2.head(15)
Out[90]:
         Permit Creation Date | Current Status Date | Issued Date | Completed Date
      0 2015-05-06
                      960
                                   187
                                          13004
                      471
                                  471
        2016-04-19
                                           12655
      3 2016-11-07
                      259
                                  253
                                          259
      4 2016-11-28
                      368
                                   368
                                          12432
                      22
      5 2017-06-14
                                           12234
      6 2017-06-30
                                          12
                      12
       7 | 2017-06-30
                                           12218
      8 2017-07-05
                                          12213
      9 2017-07-06
                                           12212
       10 2017-07-06
                                           12212
      11 2017-07-07
                                          12211
       12 | 2017-07-10
                                           12208
       14 | 2017-07-26
                                          12192
       15 | 2017-07-27
                                          12191
       16 2017-07-27
                                          12191
In [91]: #On rajoute nos données modifiées dans le dataset initial
      a = np.array(df_date2['Completed Date'].values.tolist())
      df date2['Completed Date'] = np.where(a > 5000, 9999, a).tolist()
      df_permit['Current Status Date'] = df_date2['Current Status Date']
      df permit['Issued Date'] = df_date2['Issued Date']
      df permit['Completed Date'] = df_date2['Completed Date']
      df_permit.head()
Out[91]:
                                                                First
                                                                                       Existing
                                          Current
               Permit
                      Permit
                            Street | Street |
                                     Current
                                               Issued | Completed | Construction
                                                                      Proposed
                                                                                 TIDF
        Permit
                                                                                    Construction
                Type
                     Creation
                                           Status
                            Name Suffix
                                     Status
                                                                        Units Compliance
         Type
                                                 Date
                                                        Date
                                                             Document
             Definition
                                                                                         Type
                       Date
                                            Date
                                                                Date
                                                                     2.0
             sign - erect | 05/06/2015 | Ellis
                                          960
                                                    9999
                                                            11/09/2015
                                                187
                                    expired
                                                                                    3.0
                                                                                             COI
                                                                    ... 2.0
             sign - erect | 04/19/2016 | Geary | St
                                          471
                                               471
                                                    9999
                                                                                    3.0
                                                           08/03/2017
                                    issued
                                                                                             COI
                                                                                             WO
                                    complete 259
                    11/07/2016 | Pacific | Av
                                               253
                                                    259
                                                                                    5.0
      3 8
                                                                    ... 1.0
             alterations
                                                           07/18/2017
                                                                                            (5)
             permit
      4 6
                                               368
                                                    9999
                                                                     2.0
             demolitions 11/28/2016 Market St
                                          368
                                                                                    3.0
                                                            11/20/2017
                                    issued
                                                                                             COI
             otc
      5 8
             alterations | 06/14/2017 | Indiana | St
                                               22
                                                    9999
                                                           07/06/2017
                                                                    ... 326.0
                                                                                    1.0
                                    issued
                                                                                             CO
             permit
      5 rows × 27 columns
      #One hot encoder : on convertit les données de type string en integer pour les passer dans notre réseau de neurones
      from sklearn.preprocessing import LabelEncoder,OneHotEncoder
      labelencoder_X_1 = LabelEncoder()
      df_permit_encoded = df_permit
      df_permit_encoded['Permit Type Definition'] =labelencoder_X_1.fit_transform(df_permit_encoded['Permit Type Definition'])
      df_permit_encoded['Street Name'] =labelencoder_X_1.fit_transform(df_permit_encoded['Street Name'])
      df_permit_encoded['Street Suffix'] =labelencoder_X_1.fit_transform(df_permit_encoded['Street Suffix'].astype(str))
      df_permit_encoded['Current Status'] =labelencoder_X_1.fit_transform(df_permit_encoded['Current Status'].astype(str))
      df_permit_encoded['Existing Construction Type Description'] = labelencoder_X_1.fit_transform(df_permit_encoded['Existing Construction Type Description']
      tion Type Description'].astype(str))
      df_permit_encoded['Structural Notification'] = labelencoder_X_1.fit_transform(df_permit_encoded['Structural Notification'].astype
      (str))
      df_permit_encoded['TIDF Compliance'] =labelencoder_X_1.fit_transform(df_permit_encoded['TIDF Compliance'].astype(str))
      df permit_encoded['Site Permit'] =labelencoder_X_1.fit_transform(df_permit_encoded['Site Permit'].astype(str))
      df_permit_encoded['Neighborhoods - Analysis Boundaries'] =labelencoder_X_1.fit_transform(df_permit_encoded['Neighborhoods - Anal
      ysis Boundaries'].astype(str))
      df_permit_encoded['Proposed Construction Type Description'] = labelencoder_X_1.fit_transform(df_permit_encoded['Proposed Construction Type Description']
      tion Type Description'].astype(str))
      df permit encoded['Number of Proposed Stories'] =labelencoder_X_1.fit_transform(df_permit_encoded['Number of Proposed Stories'].
      astype(str))
      df_permit_encoded['Voluntary Soft-Story Retrofit'] =labelencoder_X_1.fit_transform(df_permit_encoded['Voluntary Soft-Story Retro
      fit'].astype(str))
      df_permit_encoded['Fire Only Permit'] =labelencoder_X_1.fit_transform(df_permit_encoded['Fire Only Permit'].astype(str))
      df permit encoded['Revised Cost'] = labelencoder X 1.fit transform(df permit encoded['Revised Cost'].astype(str))
      df permit encoded['Proposed Use'] =labelencoder X 1.fit transform(df permit encoded['Proposed Use'].astype(str))
      df_permit_encoded.head()
Out[92]:
                                                               First
                                                                                     Existing
               Permit
                      Permit
                                        Current
                                              Issued | Completed | Construction
                          Street | Street | Current
                                                                               TIDF
        Permit
                                                                    Proposed
                                         Status
                                                                                   Construction
                Type
                     Creation
                           Name | Suffix |
                                                                       Units | Compliance
                                   Status
                                                      Date
                                                           Document
                                               Date
         Type
             Definition
                       Date
                                          Date
                                                               Date
                                                                                            Des
      0 4
                   05/06/2015 | 526
                                        960
                                              187
                                                          11/09/2015
                                                                    2.0
                                                                                  3.0
                                                  9999
                                                                    2.0
                   04/19/2016 | 636
                                              471
                                                          08/03/2017
                                                                                  3.0
                                        471
                                                  9999
      3 8
                                                                    1.0
                   11/07/2016 | 1176
                                        259
                                              253
                                                  259
                                                                                  5.0
                                                          07/18/2017
      4 6
                   11/28/2016 | 994
                               17
                                                                    2.0
                                        368
                                              368
                                                  9999
                                                                                  3.0
                                                          11/20/2017
      5 8
                                              22
                               17
                                        22
                                                                                  1.0
                   06/14/2017 | 781
                                                                    326.0
                                                  9999
                                                          07/06/2017
      5 rows × 27 columns
      #Probleme : la majorité des données ne mènent pas à l'obtention d'un permis. Dans notre cas, il faut avoir
      # 50% de permis validés et 50% de permis refusés pour optimiser les perf de notre réseau de neurones
      df_permit_validated = df_permit_encoded.loc[df_permit_encoded['Site Permit'] == 1]
      df_permit_refused = df_permit_encoded.loc[df_permit_encoded['Site Permit'] == 0]
      df_permit_encoded_equal = df_permit_validated.append(df_permit_refused.iloc[:3199,])
      df_permit_encoded_equal.sample(frac=1)
      df_permit_encoded_equal = df_permit_encoded_equal.drop(['Permit Creation Date'],axis=1)
      df_permit_encoded_equal = df_permit_encoded_equal.drop(['First Construction Document Date'],axis=1)
      df_permit_encoded_equal.head()
Out[93]:
                                                            Number
                                   Current
                                                                                     Existing
                Permit |
                     | Street | Street | Current |
                                        Issued Completed
                                                                   Proposed
                                                                              TIDF
                                                                                          Constr
          Permit
                                                     Structural
                                                                                  Construction
                 Type
                                   Status
                     Name Suffix
                                                 Date Notification Existing
                                                                      Units Compliance
                              Status
          Type
                                         Date
              Definition
                                                                                       Type
                                    Date
                                                                                           Desc
                                                            Stories
                                        194
       193 1
                     636
                                   194
                                             9999
                                                                   0.0
                                                            0.0
                                                                                  0.0
      583 3
                     1372
                         21
                                   762
                                        231
                                             762
                                                            2.0
                                                                   1.0
                                                                                  5.0
       767 2
                     639
                                   817
                                        161
                                             817
                                                            0.0
                                                                   1.0
                                                                                  0.0
      808 3
                     497
                                        284
                                             985
                                                            2.0
                                                                   1.0
                                   985
                                                                                  5.0
                         17
      839 3
                     32
                                        333
                                             9999
                                                            2.0
                                                                   1.0
                                                                                  5.0
                                   333
      5 rows × 25 columns
In [94]: #On normalise les données pour diminuer le temps de traitement de notre algo et augmenter ses perf
      from sklearn import preprocessing
      min_max_scaler = preprocessing.MinMaxScaler()
      np_scaled = min_max_scaler.fit_transform(df_permit_encoded_equal)
      df_normalized = pd.DataFrame(np_scaled)
      df_normalized.columns = df_permit_encoded_equal.columns
      X = df_normalized.drop(['Site Permit', 'Completed Date'],axis=1)
      Y = df_permit_encoded_equal['Site Permit']
In [95]: # Création de l'algo
      from sklearn.model_selection import train_test_split
      #On sépare le dataset en 4 dataset pour l'entrainer et le tester
      X_train, X_test, y_train, y_test = train_test_split(X, Y, test_size=0.2)
      from keras.models import Sequential
      from keras.layers import Dense, Dropout, Activation
      from keras.callbacks import EarlyStopping, ModelCheckpoint
      from keras import optimizers
      # Initialising the ANN : création des différentes couches
      classifier = Sequential()
      classifier.add(Dense(units=15, activation='relu', input_dim=23))
      classifier.add(Dropout(0.2))
      classifier.add(Dense(units=50, activation='relu'))
      classifier.add(Dropout(0.5))
      classifier.add(Dense(units=5, activation='relu'))
      classifier.add(Dropout(0.5))
      classifier.add(Dense(1, activation='sigmoid'))
      #Sélection de la méthode d'optimisation
      classifier.compile(loss='binary_crossentropy',
                optimizer=optimizers.RMSprop(lr=1e-4),
                metrics=['acc'])
      callbacks = [EarlyStopping(monitor='loss', patience=2)]
      history = classifier.fit(X_train, y_train,
                          batch_size=10,
                          epochs=50,
                          # callbacks=callbacks,
                          validation_data=(X_test, y_test))
      import matplotlib.pyplot as plt
      print(history.history.keys())
      # summarize history for loss
      plt.plot(history.history['loss'])
      plt.plot(history.history['acc'])
      plt.title('model loss')
      plt.ylabel('loss')
      plt.xlabel('epoch')
      plt.legend(['train', 'test'], loc='upper left')
      plt.show()
      scores_train = classifier.evaluate(X_train, y_train, verbose=0)
      scores_test = classifier.evaluate(X_test, y_test, verbose=0)
      print("scores_train = ", scores_train[1] * 100, '%')
      print("scores_test = ", scores_test[1] * 100, '%')
      Train on 5118 samples, validate on 1280 samples
      Epoch 1/50
      Epoch 2/50
      Epoch 3/50
      Epoch 4/50
      Epoch 5/50
      Epoch 6/50
      Epoch 7/50
      Epoch 9/50
      Epoch 10/50
      Epoch 11/50
      Epoch 13/50
      Epoch 14/50
      Epoch 15/50
      Epoch 17/50
      Epoch 18/50
      Epoch 19/50
      Epoch 21/50
      Epoch 22/50
      Epoch 23/50
      Epoch 24/50
      Epoch 25/50
      Epoch 26/50
      Epoch 27/50
      Epoch 28/50
      Epoch 29/50
      Epoch 31/50
      Epoch 32/50
      Epoch 33/50
      Epoch 35/50
      Epoch 36/50
      Epoch 37/50
      Epoch 39/50
      Epoch 40/50
      Epoch 41/50
      Epoch 43/50
      Epoch 44/50
      Epoch 45/50
      Epoch 47/50
      Epoch 48/50
      Epoch 49/50
      Epoch 50/50
      dict_keys(['val_loss', 'val_acc', 'loss', 'acc'])
                      model loss
             train
        0.9
              test
        0.8
        0.7
```

0.4 0.3 0.2 10 20 30 40 50 epoch scores\_train = 98.78858929269245 % scores\_test = 98.75 %