

mod195168les

February 9, 2023

1 Tests des modèles (cnn, rnn, bert) avec Melusine

1.0.1 Importation des packages et fonction nécessaires

```
[1]: from melusine.nlp_tools.phraser import phraser_on_body, Phraser
from melusine.utils.transformer_scheduler import TransformerScheduler
from melusine.prepare_email.manage_transfer_reply import
    ↪check_mail_begin_by_transfer
from melusine.prepare_email.manage_transfer_reply import
    ↪update_info_for_transfer_mail
from melusine.prepare_email.manage_transfer_reply import add_boolean_transfer
from melusine.prepare_email.manage_transfer_reply import add_boolean_answer
from melusine.prepare_email.body_header_extraction import extract_last_body
from melusine.prepare_email.cleaning import clean_body
from melusine.prepare_email.build_historic import build_historic
from melusine.prepare_email.mail_segmenting import structure_email
from melusine.nlp_tools.tokenizer import Tokenizer
from melusine.summarizer.keywords_generator import KeywordsGenerator
from sklearn.pipeline import Pipeline
import ast
from melusine.prepare_email.metadata_engineering import MetaExtension
from melusine.prepare_email.metadata_engineering import MetaDate
from melusine.prepare_email.metadata_engineering import MetaAttachmentType
from melusine.prepare_email.metadata_engineering import Dummifier
from melusine.nlp_tools.embedding import Embedding
from sklearn.preprocessing import LabelEncoder
from melusine.models.train import NeuralModel
from melusine.models.neural_architectures import cnn_model, bert_model,
    ↪rnn_model, transformers_model
import joblib as jl
import pandas as pd
```

C:\Users\JUFL\Anaconda3\lib\site-packages\gensim\similarities_init_.py:15:
UserWarning: The gensim.similarities.levenshtein submodule is disabled, because
the optional Levenshtein package <<https://pypi.org/project/python-Levenshtein/>>
is unavailable. Install Levenshtein (e.g. `pip install python-Levenshtein`) to
suppress this warning.

warnings.warn(msg)

1.0.2 Importation du fichier csv de tests (mails issus d'une boîte Verlingue)

```
[2]: df_emails = pd.read_csv(r"filepath", delimiter=";")
```

```
[3]: df_emails
```

```
[3]:
```

	header \	body	date \	from \	to	Cc	Cci \
0	Probl me connexion Prod et pr prod postgresql			Le Moal Mikael	DSI Infra DBA	DSI Data Factory	NaN
1	TR: package open source traitement des emails ...			L'hotellier Jean-Marie			
2	RE7_Coverlife/TALEND_BI_COVERLIFE/SID_CVLF_CHA...			DSI-suivi-production@netdmx.com			
3	DWH ADP OK - GENERATION_PRD			no-reply@verlingue.fr			
4	RE7_Coverlife/TALEND_BI_COVERLIFE/SID_CVLF_CHA...			DSI-suivi-production@netdmx.com			
..			
64	TR: Gestion de la BAL Verlingue Connect			L'hotellier Jean-Marie			
65	DWH ADP OK - VERLINGUE_PRD			no-reply@verlingue.fr			
66	DWH ADP OK - GENERATION_PRD			no-reply@verlingue.fr			
67	[MEP de Version MAJEURE 2021.V2.0] - [FIN DE R...			Fralin Thomas			
68	RE7_Coverlife/TALEND_BI_COVERLIFE/SID_CVLF_CHA...			DSI-suivi-production@netdmx.com			
0	Bonjour,\n\n \n\nLa production et la pr produc...	14/06/2021					
1	\n\n \n\nDe : Marie Leduc <mleduc@quantmetry...	14/06/2021					
2	Le traitement SID_CVLF_CHARGEMENT ne s'est pas...	14/06/2021					
3	Bonjour,\n\nLa mise disposition des donn es ...	14/06/2021					
4	Le traitement SID_CVLF_CHARGEMENT ne s'est pas...	14/06/2021					
..			
64	\n\n \n\n-----Rendez-vous d'origine-----\nDe ...	18/06/2021					
65	Bonjour,\n\nLa mise disposition des donn es ...	19/06/2021					
66	Bonjour,\n\nLa mise disposition des donn es ...	19/06/2021					
67	Information de fin de RECETTE post-MEP\n\nnde l...	19/06/2021					
68	Le traitement SID_CVLF_CHARGEMENT ne s'est pas...	19/06/2021					

```

1          Allain Marc;Fresnel Julien          NaN NaN
2          DSI Data Factory          NaN NaN
3 DSI Data Factory;Generation Statistiques;Suppo... NaN NaN
4          DSI Data Factory          NaN NaN
..          ...          ...
64          Fresnel Julien          NaN NaN
65          DSI Data Factory;jvieaud@asi.fr          NaN NaN
66 DSI Data Factory;Generation Statistiques;Suppo... NaN NaN
67 MOA Verlingue;webmoa;DSSI Verlingue;DSI Presta... DSI Management NaN
68          DSI Data Factory          NaN NaN

```

	Cat gories	Crit re de diffusion	Importance	attachment	label
0	NaN	Normal(e)	Normal(e)	NaN	production
1	NaN	Normal(e)	Normal(e)	NaN	rdv
2	NaN	Normal(e)	Normal(e)	NaN	production
3	NaN	Normal(e)	Normal(e)	NaN	datawarehouse
4	NaN	Normal(e)	Normal(e)	NaN	traitement
..
64	NaN	Normal(e)	Normal(e)	NaN	rdv
65	NaN	Normal(e)	Normal(e)	NaN	datawarehouse
66	NaN	Normal(e)	Normal(e)	NaN	datawarehouse
67	NaN	Normal(e)	Normal(e)	NaN	production
68	NaN	Normal(e)	Normal(e)	NaN	traitement

[69 rows x 12 columns]

1.0.3 Traitement sur les mails (lire notebook “Test melusine.ipynb”)

```

[4]: ManageTransformer =_
    ↪TransformerScheduler(functions_scheduler=[(check_mail_begin_by_transfer,_
    ↪None, ['is_begin_by_transfer']), (update_info_for_transfer_mail, None,_
    ↪None), (add_boolean_answer, None, ['is_answer']), (add_boolean_transfer,_
    ↪None, ['is_transfer'])])

df_emails = ManageTransformer.fit_transform(df_emails)

SegmentingTransformer =_
    ↪TransformerScheduler(functions_scheduler=[(build_historic,None,_
    ↪['structured_historic']), (structure_email, None, ['structured_body'])])

df_emails = SegmentingTransformer.fit_transform(df_emails)

Cleaning = TransformerScheduler(functions_scheduler=[(extract_last_body, None,_
    ↪['last_body']), (clean_body, None, ['clean_body'])])

df_emails = Cleaning.fit_transform(df_emails)

```

```

phraser = Phraser()
phraser.train(df_emails)

phraser.save(r'...\phraser.pickle')

PhraserTransformer = □
    ↳TransformerScheduler(functions_scheduler=[(phraser_on_body, (phraser, ), □
    ↳['clean_body'])])
df_emails = PhraserTransformer.fit_transform(df_emails)

tokenizer = Tokenizer(input_column='clean_body')
df_emails = tokenizer.fit_transform(df_emails)

keywords_generator = KeywordsGenerator(n_max_keywords=5)
df_emails = keywords_generator.fit_transform(df_emails)

pretrained_embedding = Embedding(input_column='clean_body', workers=1, □
    ↳min_count=5)
pretrained_embedding.train(df_emails)

pretrained_embedding.save(r'...\embedding.pickle')

Metadatapipeline = Pipeline([('MetaExtension', MetaExtension()), ('MetaDate', □
    ↳MetaDate()), ('MetaAttachmentType', MetaAttachmentType()), ('Dummifier', □
    ↳Dummifier())])

df_meta = Metadatapipeline.fit_transform(df_emails)

```

```

23/06 08:55 - melusine.nlp_tools.phraser - INFO -
Start training for colocation detector
23/06 08:55 - melusine.nlp_tools.phraser - INFO -
Done.

23/06 08:55 - melusine.nlp_tools.embedding - INFO - Start training for embedding
23/06 08:55 - melusine.nlp_tools.embedding - INFO - Done.

```

```
[5]: df_meta
```

```

[5]:      extension__0  extension__1  extension__2  extension__3  attachment_type__0
0                1                0                0                0                1
1                1                0                0                0                1
2                0                1                0                0                1
3                0                0                1                0                1
4                0                1                0                0                1
..                ...                ...                ...                ...
64               1                0                0                0                1

```

65	0	0	1	0	1
66	0	0	1	0	1
67	1	0	0	0	1
68	0	1	0	0	1

[69 rows x 5 columns]

```
[6]: df_emails
```

```
[6]:                                     header \
0      Problme connexion Prod et pr prod postgresql
1      TR: package open source traitement des emails ...
2      RE7_Coverlife/TALEND_BI_COVERLIFE/SID_CVLF_CHA...
3      DWH ADP OK - GENERATION_PRD
4      RE7_Coverlife/TALEND_BI_COVERLIFE/SID_CVLF_CHA...
..
64      TR: Gestion de la BAL Verlingue Connect
65      DWH ADP OK - VERLINGUE_PRD
66      DWH ADP OK - GENERATION_PRD
67      [MEP de Version MAJEURE 2021.V2.0] - [FIN DE R...
68      RE7_Coverlife/TALEND_BI_COVERLIFE/SID_CVLF_CHA...

                                     body date \
0      Bonjour,\n\n\nLa production et la pr produc... NaT
1      \n\n\nDe : Marie Leduc <mleduc@quantmetry... NaT
2      Le traitement SID_CVLF_CHARGEMENT ne s'est pas... NaT
3      Bonjour,\n\nLa mise disposition des donn es ... NaT
4      Le traitement SID_CVLF_CHARGEMENT ne s'est pas... NaT
..
64      \n\n\n-----Rendez-vous d'origine-----\nDe ... NaT
65      Bonjour,\n\nLa mise disposition des donn es ... NaT
66      Bonjour,\n\nLa mise disposition des donn es ... NaT
67      Information de fin de RECETTE post-MEP\n\nnde l... NaT
68      Le traitement SID_CVLF_CHARGEMENT ne s'est pas... NaT

                                     from \
0      Le Moal Mikael
1      L'hotellier Jean-Marie
2      DSI-suivi-production@netdmx.com
3      no-reply@verlingue.fr
4      DSI-suivi-production@netdmx.com
..
64      L'hotellier Jean-Marie
65      no-reply@verlingue.fr
66      no-reply@verlingue.fr
67      Fralin Thomas
68      DSI-suivi-production@netdmx.com
```

	to	Cc	Cci	\
0	DSI Infra DBA	DSI Data Factory	NaN	
1	Allain Marc;Fresnel Julien		NaN	NaN
2	DSI Data Factory		NaN	NaN
3	DSI Data Factory;Generation Statistiques;Suppo...		NaN	NaN
4	DSI Data Factory		NaN	NaN
..
64	Fresnel Julien		NaN	NaN
65	DSI Data Factory;jvieaud@asi.fr		NaN	NaN
66	DSI Data Factory;Generation Statistiques;Suppo...		NaN	NaN
67	MOA Verlingue;webmoa;DSSI Verlingue;DSI Presta...	DSI Management	NaN	
68	DSI Data Factory		NaN	NaN

	Cat gories	Crit re de diffusion	Importance	...	\
0	NaN	Normal(e)	Normal(e)	...	
1	NaN	Normal(e)	Normal(e)	...	
2	NaN	Normal(e)	Normal(e)	...	
3	NaN	Normal(e)	Normal(e)	...	
4	NaN	Normal(e)	Normal(e)	...	
..	
64	NaN	Normal(e)	Normal(e)	...	
65	NaN	Normal(e)	Normal(e)	...	
66	NaN	Normal(e)	Normal(e)	...	
67	NaN	Normal(e)	Normal(e)	...	
68	NaN	Normal(e)	Normal(e)	...	

	structured_body	\
0	[{'meta': {'date': None, 'from': None, 'to': N...	
1	[{'meta': {'date': None, 'from': 'Marie Leduc ...	
2	[{'meta': {'date': None, 'from': None, 'to': N...	
3	[{'meta': {'date': None, 'from': None, 'to': N...	
4	[{'meta': {'date': None, 'from': None, 'to': N...	
..	...	
64	[{'meta': {'date': None, 'from': None, 'to': N...	
65	[{'meta': {'date': None, 'from': None, 'to': N...	
66	[{'meta': {'date': None, 'from': None, 'to': N...	
67	[{'meta': {'date': None, 'from': None, 'to': N...	
68	[{'meta': {'date': None, 'from': None, 'to': N...	

	last_body	\
0	La production et la pr production PostgreSQL ...	
1	Envoy : mardi 8 juin 2021 08:37 : L'hotelli...	
2	Le traitement SID_CVLF_CHARGEMENT ne s'est pas...	
3	La mise disposition des donnees du jour dans...	
4	Le traitement SID_CVLF_CHARGEMENT ne s'est pas...	
..	...	

```

64          -----Rendez-vous d'origine-----
65 La mise disposition des donnees du jour dans...
66 La mise disposition des donnees du jour dans...
67 Information de fin de RECETTE post-MEP de la V...
68 Le traitement SID_CVLF_CHARGEMENT ne s'est pas...

                                clean_body \
0  la production et la prproduction postgresql se...
1  envoy : flag_date_ flag_time_ : l'hotellie...
2  le traitement sid_cvlf_chargement ne s'est pas...
3  la mise disposition des donnees du jour dans le...
4  le traitement sid_cvlf_chargement ne s'est pas...
..
64          rendez-vous d'origine
65 la mise disposition des donnees du jour dans le...
66 la mise disposition des donnees du jour dans le...
67 information de fin de recette post-mep de la v...
68 le traitement sid_cvlf_chargement ne s'est pas...

                                tokens \
0  [production, prproduction, postgresql, semble,...
1  [envoy, flag_date_, flag_time_, hotellier, fla...
2  [traitement, sid_cvlf_chargement, termin, corr...
3  [mise, disposition, donnees, jour, datawarehous...
4  [traitement, sid_cvlf_chargement, termin, corr...
..
64          [rendez-vous, origine]
65 [mise, disposition, donnees, jour, datawarehous...
66 [mise, disposition, donnees, jour, datawarehous...
67 [information, fin, recette, post-mep, version,...
68 [traitement, sid_cvlf_chargement, termin, corr...

                                keywords extension hour min \
0  [prproduction, flag_date_, vers, flag_time_, c...      0 NaN NaN
1  [flag_date_, flag_time_, hotellier, flag_name_...      0 NaN NaN
2  [traitement, sid_cvlf_chargement, suivi, produ...      1 NaN NaN
3  [atlas, nombre, flag_cp_, flag_date_, pouvez]          2 NaN NaN
4  [traitement, sid_cvlf_chargement, suivi, produ...      1 NaN NaN
..
64          [origine]                                     ... ..
65 [nombre, flag_cp_, flag_date_, gnration, pouvez]        2 NaN NaN
66 [nombre, flag_phone_, flag_cp_, flag_date_, po...      2 NaN NaN
67 [flag_time_, tnr, flag_name_, iard, flag_url_]          0 NaN NaN
68 [traitement, sid_cvlf_chargement, suivi, produ...      1 NaN NaN

dayofweek attachment_type
0          NaN          [0]

```

1	NaN	[0]
2	NaN	[0]
3	NaN	[0]
4	NaN	[0]
..
64	NaN	[0]
65	NaN	[0]
66	NaN	[0]
67	NaN	[0]
68	NaN	[0]

[69 rows x 26 columns]

1.0.4 Modèles NLP

Dans le package *melusine* il existe plusieurs modèles nlp :

- Le modèle **RNN** : il est utilisé traditionnellement pour les données textuelles.
- Le modèle **CNN** : modèle un peu moins précis que le rnn, il est cependant plus rapide et demande encore moins de ressources mémoire
- Le **Transformers model** : modèle aussi précis que le rnn, plus rapide que ce dernier et qui demande autant de ressource mémoire
- Le modèle **CamemBert** : architecture bert
- Le modèle **FlauBert** : architecture bert

Voici comment appeler chaque modèle :

On utilise la méthode *LabelEncoder* qui va transformer des *str* en *int*

```
[7]: X = pd.concat([df_emails['clean_body'], df_meta], axis=1)
y = df_emails['label']

le = LabelEncoder()
y = le.fit_transform(y)

nn_model = NeuralModel(architecture_function= cnn_model,
    ↳ pretrained_embedding=pretrained_embedding, text_input_column=
    ↳ "clean_body", meta_input_list=['extension', 'dayofweek', 'hour', 'min',
    ↳ 'attachment_type'], n_epochs=10)
nn_model2 = NeuralModel(architecture_function= rnn_model,
    ↳ pretrained_embedding=pretrained_embedding, text_input_column=
    ↳ "clean_body", meta_input_list=['extension', 'dayofweek', 'hour', 'min',
    ↳ 'attachment_type'], n_epochs=10)
nn_model3 = NeuralModel(architecture_function= rnn_model,
    ↳ pretrained_embedding=pretrained_embedding, text_input_column=
    ↳ "clean_body", meta_input_list=['extension', 'dayofweek', 'hour', 'min',
    ↳ 'attachment_type'], n_epochs=10)
```



```

nn_model4 = NeuralModel(architecture_function= bert_model,
    ↳pretrained_embedding=pretrained_embedding, text_input_column=
    ↳="clean_body",meta_input_list=['extension', 'dayofweek','hour', 'min',
    ↳'attachment_type'], n_epochs=1, bert_tokenizer='jplu/tf-camembert-base',
    ↳bert_model='jplu/tf-camembert-base')
nn_model5 = NeuralModel(architecture_function= bert_model,
    ↳pretrained_embedding=pretrained_embedding, text_input_column=
    ↳="clean_body",meta_input_list=['extension', 'dayofweek','hour', 'min',
    ↳'attachment_type'], n_epochs=1, bert_tokenizer='jplu/
    ↳tf-flaubert-base-cased', bert_model='jplu/tf-flaubert-base-cased')

```

On effectue l'entrainement avec la méthode *fit*

```

[8]: print('-----cnn_model-----')
nn_model.fit(X,y)
print('-----rnn_model-----')
nn_model2.fit(X,y)
print('-----transformers_model-----')
nn_model3.fit(X,y)
print('-----CamemBert_model-----')
nn_model4.fit(X,y)
print('-----FlauBert_model-----')
nn_model5.fit(X,y)

```

```

-----cnn_model-----
Epoch 1/10
1/1 [=====] - 0s 995us/step - loss: 2.3005 - accuracy:
0.1014
Epoch 2/10
1/1 [=====] - 0s 0s/step - loss: 2.1340 - accuracy:
0.1739
Epoch 3/10
1/1 [=====] - 0s 0s/step - loss: 2.1062 - accuracy:
0.1304
Epoch 4/10
1/1 [=====] - 0s 999us/step - loss: 2.1275 - accuracy:
0.2029
Epoch 5/10
1/1 [=====] - 0s 0s/step - loss: 2.0542 - accuracy:
0.2464
Epoch 6/10
1/1 [=====] - 0s 0s/step - loss: 2.0007 - accuracy:
0.1739
Epoch 7/10
1/1 [=====] - 0s 1000us/step - loss: 1.9962 - accuracy:
0.2899
Epoch 8/10
1/1 [=====] - 0s 1000us/step - loss: 1.9542 - accuracy:

```

```

0.2609
Epoch 9/10
1/1 [=====] - 0s 996us/step - loss: 1.9863 - accuracy:
0.3043
Epoch 10/10
1/1 [=====] - 0s 997us/step - loss: 1.9930 - accuracy:
0.2464
-----rnn_model-----
Epoch 1/10
1/1 [=====] - 0s 16ms/step - loss: 2.2060 - accuracy:
0.1304
Epoch 2/10
1/1 [=====] - 0s 999us/step - loss: 2.1811 - accuracy:
0.1594
Epoch 3/10
1/1 [=====] - 0s 0s/step - loss: 2.1582 - accuracy:
0.2754
Epoch 4/10
1/1 [=====] - 0s 2ms/step - loss: 2.1385 - accuracy:
0.2319
Epoch 5/10
1/1 [=====] - 0s 0s/step - loss: 2.0998 - accuracy:
0.2464
Epoch 6/10
1/1 [=====] - 0s 998us/step - loss: 2.0814 - accuracy:
0.2609
Epoch 7/10
1/1 [=====] - 0s 16ms/step - loss: 2.0511 - accuracy:
0.2464
Epoch 8/10
1/1 [=====] - 0s 0s/step - loss: 2.0728 - accuracy:
0.2464
Epoch 9/10
1/1 [=====] - 0s 0s/step - loss: 1.9876 - accuracy:
0.2609
Epoch 10/10
1/1 [=====] - 0s 0s/step - loss: 2.0279 - accuracy:
0.2319
-----transformers_model-----
Epoch 1/10
1/1 [=====] - 0s 18ms/step - loss: 2.2013 - accuracy:
0.1304
Epoch 2/10
1/1 [=====] - 0s 0s/step - loss: 2.1823 - accuracy:
0.0870
Epoch 3/10
1/1 [=====] - 0s 998us/step - loss: 2.1478 - accuracy:
0.2609

```

```

Epoch 4/10
1/1 [=====] - 0s 0s/step - loss: 2.1215 - accuracy:
0.2174
Epoch 5/10
1/1 [=====] - 0s 0s/step - loss: 2.0944 - accuracy:
0.2754
Epoch 6/10
1/1 [=====] - 0s 0s/step - loss: 2.0778 - accuracy:
0.2609
Epoch 7/10
1/1 [=====] - 0s 1ms/step - loss: 2.0416 - accuracy:
0.2319
Epoch 8/10
1/1 [=====] - 0s 0s/step - loss: 2.0390 - accuracy:
0.2464
Epoch 9/10
1/1 [=====] - 0s 2ms/step - loss: 1.9538 - accuracy:
0.2754
Epoch 10/10
1/1 [=====] - 0s 0s/step - loss: 2.0039 - accuracy:
0.2899
-----CamemBert_model-----

```

Some layers from the model checkpoint at jplu/tf-camembert-base were not used when initializing TFCamembertModel: ['lm_head']

- This IS expected if you are initializing TFCamembertModel from the checkpoint of a model trained on another task or with another architecture (e.g. initializing a BertForSequenceClassification model from a BertForPretraining model).

- This IS NOT expected if you are initializing TFCamembertModel from the checkpoint of a model that you expect to be exactly identical (initializing a BertForSequenceClassification model from a BertForSequenceClassification model). All the layers of TFCamembertModel were initialized from the model checkpoint at jplu/tf-camembert-base.

If your task is similar to the task the model of the checkpoint was trained on, you can already use TFCamembertModel for predictions without further training.

```

1/1 [=====] - 0s 131ms/step - loss: 2.2079 - accuracy:
0.0725
-----FlauBert_model-----

```

Some layers from the model checkpoint at jplu/tf-flaubert-base-cased were not used when initializing TFFlaubertModel: ['pred_layer_.proj']

- This IS expected if you are initializing TFFlaubertModel from the checkpoint of a model trained on another task or with another architecture (e.g. initializing a BertForSequenceClassification model from a BertForPretraining model).

- This IS NOT expected if you are initializing TFFlaubertModel from the checkpoint of a model that you expect to be exactly identical (initializing a

BertForSequenceClassification model from a BertForSequenceClassification model). All the layers of TFFlaubertModel were initialized from the model checkpoint at jplu/tf-flaubert-base-cased.

If your task is similar to the task the model of the checkpoint was trained on, you can already use TFFlaubertModel for predictions without further training.

```
1/1 [=====] - 0s 155ms/step - loss: 2.5251 - accuracy: 0.1594
```

Maintenant que les modèles sont entraînés nous allons les tester sur le même jeu de données afin de voir si les prédictions sont correctes. (Et histoire de montrer le fonctionnement)

1.0.5 Tests

Utilisons le même jeu de données mais non-labellisées :

```
[9]: df_emails_tests = pd.read_csv(r"...testtrain2.CSV", delimiter=";")
```

```
[10]: df_emails_tests
```

```
[10]:
                                     header \
0      Problme connexion Prod et pr prod postgresql
1      TR: package open source traitement des emails ...
2      RE7_Coverlife/TALEND_BI_COVERLIFE/SID_CVLF_CHA...
3                                     DWH ADP OK - GENERATION_PRD
4      RE7_Coverlife/TALEND_BI_COVERLIFE/SID_CVLF_CHA...
..                                     ...
64      TR: Gestion de la BAL Verlingue Connect
65                                     DWH ADP OK - VERLINGUE_PRD
66                                     DWH ADP OK - GENERATION_PRD
67      [MEP de Version MAJEURE 2021.V2.0] - [FIN DE R...
68      RE7_Coverlife/TALEND_BI_COVERLIFE/SID_CVLF_CHA...

                                     body      date \
0      Bonjour,\n\n \n\nLa production et la pr produc... 14/06/2021
1      \n\n \n\nDe : Marie Leduc <mleduc@quantmetry... 14/06/2021
2      Le traitement SID_CVLF_CHARGEMENT ne s'est pas... 14/06/2021
3      Bonjour,\n\nLa mise disposition des donn es ... 14/06/2021
4      Le traitement SID_CVLF_CHARGEMENT ne s'est pas... 14/06/2021
..                                     ...
64      \n\n \n\n-----Rendez-vous d'origine-----\nDe ... 18/06/2021
65      Bonjour,\n\nLa mise disposition des donn es ... 19/06/2021
66      Bonjour,\n\nLa mise disposition des donn es ... 19/06/2021
67      Information de fin de RECETTE post-MEP\n\nnde l... 19/06/2021
68      Le traitement SID_CVLF_CHARGEMENT ne s'est pas... 19/06/2021

                                     from \
0                                     Le Moal Mikael
1                                     L'hotelier Jean-Marie
```

```

2 DSI-suivi-production@netdmx.com
3     no-reply@verlingue.fr
4 DSI-suivi-production@netdmx.com
..
64 L'hotellier Jean-Marie
65     no-reply@verlingue.fr
66     no-reply@verlingue.fr
67     Fralin Thomas
68 DSI-suivi-production@netdmx.com

```

	to	Cc	Cci	\
0	DSI Infra DBA	DSI Data Factory	NaN	
1	Allain Marc;Fresnel Julien		NaN	NaN
2	DSI Data Factory		NaN	NaN
3	DSI Data Factory;Generation Statistiques;Suppo...		NaN	NaN
4	DSI Data Factory		NaN	NaN
..
64	Fresnel Julien		NaN	NaN
65	DSI Data Factory;jvieaud@asi.fr		NaN	NaN
66	DSI Data Factory;Generation Statistiques;Suppo...		NaN	NaN
67	MOA Verlingue;webmoa;DSSI Verlingue;DSI Presta...	DSI Management	NaN	
68	DSI Data Factory		NaN	NaN

	Cat gories	Crit re de diffusion	Importance	attachment
0	NaN	Normal(e)	Normal(e)	NaN
1	NaN	Normal(e)	Normal(e)	NaN
2	NaN	Normal(e)	Normal(e)	NaN
3	NaN	Normal(e)	Normal(e)	NaN
4	NaN	Normal(e)	Normal(e)	NaN
..
64	NaN	Normal(e)	Normal(e)	NaN
65	NaN	Normal(e)	Normal(e)	NaN
66	NaN	Normal(e)	Normal(e)	NaN
67	NaN	Normal(e)	Normal(e)	NaN
68	NaN	Normal(e)	Normal(e)	NaN

[69 rows x 11 columns]

On réalise les mêmes traitements que d'habitude :

```

[11]: ManageTransformer = TransformerScheduler(functions_scheduler=[(check_mail_begin_by_transfer,
↳None, ['is_begin_by_transfer']), (update_info_for_transfer_mail, None,
↳None), (add_boolean_answer, None, ['is_answer']), (add_boolean_transfer,
↳None, ['is_transfer'])])

df_emails_tests = ManageTransformer.fit_transform(df_emails_tests)

```

```

SegmentingTransformer = _
    ↪ TransformerScheduler(functions_scheduler=[(build_historic, None, _
    ↪ ['structured_historic']), (structure_email, None, ['structured_body'])])

df_emails_tests = SegmentingTransformer.fit_transform(df_emails_tests)

Cleaning = TransformerScheduler(functions_scheduler=[(extract_last_body, None, _
    ↪ ['last_body']), (clean_body, None, ['clean_body'])])

df_emails_tests = Cleaning.fit_transform(df_emails_tests)

tokenizer = Tokenizer(input_column='clean_body')
df_emails_tests = tokenizer.fit_transform(df_emails_tests)

Metadatapipeline = Pipeline([('MetaExtension', MetaExtension()), ('MetaDate', _
    ↪ MetaDate()), ('MetaAttachmentType', MetaAttachmentType()), ('Dummifier', _
    ↪ Dummifier())])

df_meta_tests = Metadatapipeline.fit_transform(df_emails_tests)

```

```
[12]: df_meta_tests
```

```

[12]:
   extension__0  extension__1  extension__2  extension__3  attachment_type__0
0              1            0              0              0                  1
1              1            0              0              0                  1
2              0            1              0              0                  1
3              0            0              1              0                  1
4              0            1              0              0                  1
..            ...          ...          ...          ...          ...
64              1            0              0              0                  1
65              0            0              1              0                  1
66              0            0              1              0                  1
67              1            0              0              0                  1
68              0            1              0              0                  1

```

[69 rows x 5 columns]

```
[13]: df_emails_tests
```

```

[13]:
   header \
0      Problme connexion Prod et pr prod postgresql
1      TR: package open source traitement des emails ...
2      RE7_Coverlife/TALEND_BI_COVERLIFE/SID_CVLF_CHA...
3      DWH ADP OK - GENERATION_PRD
4      RE7_Coverlife/TALEND_BI_COVERLIFE/SID_CVLF_CHA...
..      ...

```

64 TR: Gestion de la BAL Verlingue Connect
 65 DWH ADP OK - VERLINGUE_PRD
 66 DWH ADP OK - GENERATION_PRD
 67 [MEP de Version MAJEURE 2021.V2.0] - [FIN DE R...
 68 RE7_Coverlife/TALEND_BI_COVERLIFE/SID_CVLF_CHA...

body date \
 0 Bonjour,\n\n\nLa production et la pr produc... NaT
 1 \n\n\nDe : Marie Leduc <mleduc@quantmetry... NaT
 2 Le traitement SID_CVLF_CHARGEMENT ne s'est pas... NaT
 3 Bonjour,\n\nLa mise disposition des donn es ... NaT
 4 Le traitement SID_CVLF_CHARGEMENT ne s'est pas... NaT

 64 \n\n\n-----Rendez-vous d'origine-----\nDe ... NaT
 65 Bonjour,\n\nLa mise disposition des donn es ... NaT
 66 Bonjour,\n\nLa mise disposition des donn es ... NaT
 67 Information de fin de RECETTE post-MEP\n\nnde l... NaT
 68 Le traitement SID_CVLF_CHARGEMENT ne s'est pas... NaT

from \
 0 Le Moal Mikael
 1 L'hotellier Jean-Marie
 2 DSI-suivi-production@netdmx.com
 3 no-reply@verlingue.fr
 4 DSI-suivi-production@netdmx.com

 64 L'hotellier Jean-Marie
 65 no-reply@verlingue.fr
 66 no-reply@verlingue.fr
 67 Fralin Thomas
 68 DSI-suivi-production@netdmx.com

	to	Cc	Cci	\
0	DSI Infra DBA	DSI Data Factory	NaN	
1	Allain Marc;Fresnel Julien		NaN	NaN
2	DSI Data Factory		NaN	NaN
3	DSI Data Factory;Generation Statistiques;Suppo...		NaN	NaN
4	DSI Data Factory		NaN	NaN
..
64	Fresnel Julien		NaN	NaN
65	DSI Data Factory;jvieaud@asi.fr		NaN	NaN
66	DSI Data Factory;Generation Statistiques;Suppo...		NaN	NaN
67	MOA Verlingue;webmoa;DSSI Verlingue;DSI Presta...	DSI Management	NaN	
68	DSI Data Factory		NaN	NaN

Cat gories Crit re de diffusion Importance ... \
 0 NaN Normal(e) Normal(e) ...

1	NaN	Normal(e)	Normal(e)	...
2	NaN	Normal(e)	Normal(e)	...
3	NaN	Normal(e)	Normal(e)	...
4	NaN	Normal(e)	Normal(e)	...
..
64	NaN	Normal(e)	Normal(e)	...
65	NaN	Normal(e)	Normal(e)	...
66	NaN	Normal(e)	Normal(e)	...
67	NaN	Normal(e)	Normal(e)	...
68	NaN	Normal(e)	Normal(e)	...

```

                                structured_historic \
0  [{'text': 'Bonjour,
```

La production et la pr...

```

1  [{'text': 'Envoy : mardi 8 juin 2021 08:37
...
2  [{'text': 'Le traitement SID_CVLF_CHARGEMENT n...
3  [{'text': 'Bonjour,
```

La mise disposition des...

```

4  [{'text': 'Le traitement SID_CVLF_CHARGEMENT n...
..
64 [{'text': '
...
```

-----Rendez-vous d'origine---

```

65 [{'text': 'Bonjour,
```

La mise disposition des...

```

66 [{'text': 'Bonjour,
```

La mise disposition des...

```

67 [{'text': 'Information de fin de RECETTE post-...
68 [{'text': 'Le traitement SID_CVLF_CHARGEMENT n...
```

```

                                structured_body \
0  [{'meta': {'date': None, 'from': None, 'to': N...
1  [{'meta': {'date': None, 'from': 'Marie Leduc ...
2  [{'meta': {'date': None, 'from': None, 'to': N...
3  [{'meta': {'date': None, 'from': None, 'to': N...
4  [{'meta': {'date': None, 'from': None, 'to': N...
..
64 [{'meta': {'date': None, 'from': None, 'to': N...
```



```

65 [{'meta': {'date': None, 'from': None, 'to': N...
66 [{'meta': {'date': None, 'from': None, 'to': N...
67 [{'meta': {'date': None, 'from': None, 'to': N...
68 [{'meta': {'date': None, 'from': None, 'to': N...

```

```

                                last_body \
0   La production et la pr production PostgreSQL ...
1   Envoy : mardi 8 juin 2021 08:37 : L'hotelli...
2   Le traitement SID_CVLF_CHARGEMENT ne s'est pas...
3   La mise disposition des donnees du jour dans...
4   Le traitement SID_CVLF_CHARGEMENT ne s'est pas...
..
64   -----Rendez-vous d'origine-----
65   La mise disposition des donnees du jour dans...
66   La mise disposition des donnees du jour dans...
67   Information de fin de RECETTE post-MEP de la V...
68   Le traitement SID_CVLF_CHARGEMENT ne s'est pas...

```

```

                                clean_body \
0   la production et la prproduction postgresql se...
1   envoy : flag_date_ flag_time_ : l'hotellie...
2   le traitement sid_cvlf_chargement ne s'est pas...
3   la mise disposition des donnees du jour dans le...
4   le traitement sid_cvlf_chargement ne s'est pas...
..
64   rendez-vous d'origine
65   la mise disposition des donnees du jour dans le...
66   la mise disposition des donnees du jour dans le...
67   information de fin de recette post-mep de la v...
68   le traitement sid_cvlf_chargement ne s'est pas...

```

```

                                tokens extension hour min \
0   [production, prproduction, postgresql, semble,...      0 NaN NaN
1   [envoy, flag_date_, flag_time_, hotellier, fla...      0 NaN NaN
2   [traitement, sid_cvlf_chargement, termin, corr...      1 NaN NaN
3   [mise, disposition, donnees, jour, datawarehous...    2 NaN NaN
4   [traitement, sid_cvlf_chargement, termin, corr...      1 NaN NaN
..
64   [rendez-vous, origine]                                ... ..
65   [mise, disposition, donnees, jour, datawarehous...    2 NaN NaN
66   [mise, disposition, donnees, jour, datawarehous...    2 NaN NaN
67   [information, fin, recette, post-mep, version,...      0 NaN NaN
68   [traitement, sid_cvlf_chargement, termin, corr...      1 NaN NaN

```

```

dayofweek  attachment_type
0          NaN             [0]
1          NaN             [0]

```

```

2      NaN      [0]
3      NaN      [0]
4      NaN      [0]
..     ...      ...
64     NaN      [0]
65     NaN      [0]
66     NaN      [0]
67     NaN      [0]
68     NaN      [0]

```

[69 rows x 24 columns]

On réalise les mêmes concaténations que précédemment mais avec *df_emails_tests* et *df_meta_tests*

```
[14]: X_tests = pd.concat([df_emails_tests['clean_body'], df_meta_tests], axis=1)
```

```
[15]: X_tests
```

```
[15]:
```

	clean_body	extension__0	\
0	la production et la prproduction postgresql se...	1	
1	envoy : flag_date_ flag_time_ : l'hotellie...	1	
2	le traitement sid_cvlf_chargement ne s'est pas...	0	
3	la mise disposition des donnees du jour dans le...	0	
4	le traitement sid_cvlf_chargement ne s'est pas...	0	
..	
64	rendez-vous d'origine	1	
65	la mise disposition des donnees du jour dans le...	0	
66	la mise disposition des donnees du jour dans le...	0	
67	information de fin de recette post-mep de la v...	1	
68	le traitement sid_cvlf_chargement ne s'est pas...	0	

	extension__1	extension__2	extension__3	attachment_type__0
0	0	0	0	1
1	0	0	0	1
2	1	0	0	1
3	0	1	0	1
4	1	0	0	1
..
64	0	0	0	1
65	0	1	0	1
66	0	1	0	1
67	0	0	0	1
68	1	0	0	1

[69 rows x 6 columns]

Prédictions On réalise les prédictions :

```
[16]: yr = nn_model.predict(X_tests)
      yr2 = nn_model2.predict(X_tests)
      yr3 = nn_model3.predict(X_tests)
      yr4 = nn_model4.predict(X_tests)
      yr5 = nn_model5.predict(X_tests)
```

WARNING:tensorflow:5 out of the last 13 calls to <function Model.make_predict_function.<locals>.predict_function at 0x000001FED68B9820> triggered tf.function retracing. Tracing is expensive and the excessive number of tracings could be due to (1) creating @tf.function repeatedly in a loop, (2) passing tensors with different shapes, (3) passing Python objects instead of tensors. For (1), please define your @tf.function outside of the loop. For (2), @tf.function has experimental_relax_shapes=True option that relaxes argument shapes that can avoid unnecessary retracing. For (3), please refer to https://www.tensorflow.org/tutorials/customization/performance#python_or_tensor_args and https://www.tensorflow.org/api_docs/python/tf/function for more details.

On applique *le.inverse_transform* pour changer les *int* en *str*

```
[17]: yr = le.inverse_transform(yr)
      yr2 = le.inverse_transform(yr2)
      yr3 = le.inverse_transform(yr3)
      yr4 = le.inverse_transform(yr4)
      yr5 = le.inverse_transform(yr5)
```

voici les résultats de chaque *yr* :

```
[18]: print("yr : ", yr)
      print("\n")
      print("yr2 : ", yr2)
      print("\n")
      print("yr3 : ", yr3)
      print("\n")
      print("yr4 : ", yr4)
      print("\n")
      print("yr5 : ", yr5)
```

```
yr : ['production' 'production' 'production' 'datawarehouse' 'production'
      'datawarehouse' 'production' 'production' 'production' 'production'
      'production' 'production' 'production' 'production' 'production'
      'production' 'datawarehouse' 'production' 'datawarehouse' 'production'
      'production' 'production' 'production' 'production' 'production'
      'production' 'production' 'production' 'production' 'production'
      'production' 'production' 'production' 'production' 'production'
      'production' 'production' 'production' 'production' 'production'
      'datawarehouse' 'datawarehouse' 'production' 'production' 'production'
      'production' 'production' 'production' 'production' 'production']
```



```
'assistance' 'production' 'production' 'assistance' 'production'
'assistance' 'production' 'assistance' 'production' 'production'
'production' 'assistance' 'communication' 'assistance' 'production'
'production' 'production' 'production' 'production']
```

```
yr5 : ['production' 'production' 'production' 'rdv' 'production' 'traitement'
'traitement' 'assistance' 'assistance' 'personnel' 'production'
'production' 'rdv' 'personnel' 'production' 'production' 'traitement'
'datawarehouse' 'traitement' 'production' 'rdv' 'production' 'rdv'
'traitement' 'traitement' 'production' 'rdv' 'production' 'production'
'personnel' 'personnel' 'personnel' 'production' 'production' 'personnel'
'personnel' 'production' 'traitement' 'production' 'personnel'
'traitement' 'traitement' 'personnel' 'assistance' 'production'
'personnel' 'rdv' 'production' 'assistance' 'personnel' 'personnel'
'production' 'assistance' 'personnel' 'production' 'production'
'production' 'datawarehouse' 'production' 'production' 'production'
'assistance' 'rdv' 'datawarehouse' 'rdv' 'traitement' 'traitement' 'rdv'
'production']
```

On remarque que les prédictions se sont faite, néanmoins elles ne sont pas exactes. Tentons de se rapprocher des bons résultats.

1.0.6 Bonus (80 époques pour le cnn_model)

```
[19]: nn_model6 = NeuralModel(architecture_function= cnn_model,
↳ pretrained_embedding=pretrained_embedding, text_input_column=
↳ "clean_body", meta_input_list=['extension', 'dayofweek', 'hour', 'min',
↳ 'attachment_type'], n_epochs=80)

nn_model6.fit(X,y)
```

Epoch 1/80

```
1/1 [=====] - 0s 0s/step - loss: 2.5261 - accuracy:
0.1304
```

Epoch 2/80

```
1/1 [=====] - 0s 998us/step - loss: 2.2406 - accuracy:
0.1014
```

Epoch 3/80

```
1/1 [=====] - 0s 996us/step - loss: 2.1036 - accuracy:
0.2174
```

Epoch 4/80

```
1/1 [=====] - 0s 2ms/step - loss: 2.0853 - accuracy:
0.2319
```

Epoch 5/80

```
1/1 [=====] - 0s 0s/step - loss: 2.0786 - accuracy:
0.2029
```

Epoch 6/80

```

1/1 [=====] - 0s 2ms/step - loss: 2.1691 - accuracy:
0.1884
Epoch 7/80
1/1 [=====] - 0s 2ms/step - loss: 2.0386 - accuracy:
0.2464
Epoch 8/80
1/1 [=====] - 0s 998us/step - loss: 2.0685 - accuracy:
0.2029
Epoch 9/80
1/1 [=====] - 0s 997us/step - loss: 2.0152 - accuracy:
0.2464
Epoch 10/80
1/1 [=====] - 0s 997us/step - loss: 1.9660 - accuracy:
0.2899
Epoch 11/80
1/1 [=====] - 0s 997us/step - loss: 1.8782 - accuracy:
0.3188
Epoch 12/80
1/1 [=====] - 0s 2ms/step - loss: 1.9488 - accuracy:
0.2464
Epoch 13/80
1/1 [=====] - 0s 997us/step - loss: 1.8864 - accuracy:
0.3333
Epoch 14/80
1/1 [=====] - 0s 2ms/step - loss: 1.7649 - accuracy:
0.3333
Epoch 15/80
1/1 [=====] - 0s 998us/step - loss: 1.8627 - accuracy:
0.2464
Epoch 16/80
1/1 [=====] - 0s 0s/step - loss: 1.7360 - accuracy:
0.3478
Epoch 17/80
1/1 [=====] - 0s 984us/step - loss: 1.7354 - accuracy:
0.4058
Epoch 18/80
1/1 [=====] - 0s 996us/step - loss: 1.6523 - accuracy:
0.4203
Epoch 19/80
1/1 [=====] - 0s 3ms/step - loss: 1.6751 - accuracy:
0.3478
Epoch 20/80
1/1 [=====] - 0s 998us/step - loss: 1.6273 - accuracy:
0.4203
Epoch 21/80
1/1 [=====] - 0s 0s/step - loss: 1.5143 - accuracy:
0.4058
Epoch 22/80

```

```

1/1 [=====] - 0s 996us/step - loss: 1.5104 - accuracy:
0.4348
Epoch 23/80
1/1 [=====] - 0s 996us/step - loss: 1.4416 - accuracy:
0.5217
Epoch 24/80
1/1 [=====] - 0s 2ms/step - loss: 1.4754 - accuracy:
0.3913
Epoch 25/80
1/1 [=====] - 0s 997us/step - loss: 1.3680 - accuracy:
0.4783
Epoch 26/80
1/1 [=====] - 0s 2ms/step - loss: 1.4265 - accuracy:
0.4783
Epoch 27/80
1/1 [=====] - 0s 2ms/step - loss: 1.2338 - accuracy:
0.5507
Epoch 28/80
1/1 [=====] - 0s 996us/step - loss: 1.3989 - accuracy:
0.4783
Epoch 29/80
1/1 [=====] - 0s 998us/step - loss: 1.3283 - accuracy:
0.4928
Epoch 30/80
1/1 [=====] - 0s 997us/step - loss: 1.2582 - accuracy:
0.5072
Epoch 31/80
1/1 [=====] - 0s 997us/step - loss: 1.2549 - accuracy:
0.5072
Epoch 32/80
1/1 [=====] - 0s 997us/step - loss: 1.2565 - accuracy:
0.4783
Epoch 33/80
1/1 [=====] - 0s 982us/step - loss: 1.2467 - accuracy:
0.4638
Epoch 34/80
1/1 [=====] - 0s 2ms/step - loss: 1.1538 - accuracy:
0.5362
Epoch 35/80
1/1 [=====] - 0s 0s/step - loss: 1.2498 - accuracy:
0.5072
Epoch 36/80
1/1 [=====] - 0s 994us/step - loss: 1.1480 - accuracy:
0.5507
Epoch 37/80
1/1 [=====] - 0s 998us/step - loss: 1.1061 - accuracy:
0.5507
Epoch 38/80

```

```

1/1 [=====] - 0s 1ms/step - loss: 1.0604 - accuracy:
0.6087
Epoch 39/80
1/1 [=====] - 0s 994us/step - loss: 1.0697 - accuracy:
0.5797
Epoch 40/80
1/1 [=====] - 0s 997us/step - loss: 0.9948 - accuracy:
0.6377
Epoch 41/80
1/1 [=====] - 0s 998us/step - loss: 1.0372 - accuracy:
0.6377
Epoch 42/80
1/1 [=====] - 0s 997us/step - loss: 0.9890 - accuracy:
0.6232
Epoch 43/80
1/1 [=====] - 0s 997us/step - loss: 0.9734 - accuracy:
0.6087
Epoch 44/80
1/1 [=====] - 0s 998us/step - loss: 1.1132 - accuracy:
0.5362
Epoch 45/80
1/1 [=====] - 0s 996us/step - loss: 0.8346 - accuracy:
0.6522
Epoch 46/80
1/1 [=====] - 0s 998us/step - loss: 0.8594 - accuracy:
0.6812
Epoch 47/80
1/1 [=====] - 0s 997us/step - loss: 0.7761 - accuracy:
0.6667
Epoch 48/80
1/1 [=====] - 0s 2ms/step - loss: 0.7963 - accuracy:
0.7101
Epoch 49/80
1/1 [=====] - 0s 998us/step - loss: 0.7737 - accuracy:
0.7246
Epoch 50/80
1/1 [=====] - 0s 995us/step - loss: 0.7688 - accuracy:
0.6667
Epoch 51/80
1/1 [=====] - 0s 993us/step - loss: 0.5797 - accuracy:
0.7536
Epoch 52/80
1/1 [=====] - 0s 993us/step - loss: 0.7630 - accuracy:
0.7826
Epoch 53/80
1/1 [=====] - 0s 0s/step - loss: 0.5748 - accuracy:
0.8116
Epoch 54/80

```


1/1 [=====] - 0s 2ms/step - loss: 0.5866 - accuracy:
 0.7971
 Epoch 55/80
 1/1 [=====] - 0s 0s/step - loss: 0.7082 - accuracy:
 0.7101
 Epoch 56/80
 1/1 [=====] - 0s 996us/step - loss: 0.4812 - accuracy:
 0.7826
 Epoch 57/80
 1/1 [=====] - 0s 996us/step - loss: 0.3634 - accuracy:
 0.8986
 Epoch 58/80
 1/1 [=====] - 0s 997us/step - loss: 0.4810 - accuracy:
 0.8551
 Epoch 59/80
 1/1 [=====] - 0s 997us/step - loss: 0.4985 - accuracy:
 0.8551
 Epoch 60/80
 1/1 [=====] - 0s 996us/step - loss: 0.3991 - accuracy:
 0.8261
 Epoch 61/80
 1/1 [=====] - 0s 0s/step - loss: 0.3308 - accuracy:
 0.8841
 Epoch 62/80
 1/1 [=====] - 0s 998us/step - loss: 0.3353 - accuracy:
 0.8551
 Epoch 63/80
 1/1 [=====] - 0s 998us/step - loss: 0.2650 - accuracy:
 0.9275
 Epoch 64/80
 1/1 [=====] - 0s 997us/step - loss: 0.2744 - accuracy:
 0.8986
 Epoch 65/80
 1/1 [=====] - 0s 997us/step - loss: 0.2719 - accuracy:
 0.8841
 Epoch 66/80
 1/1 [=====] - 0s 2ms/step - loss: 0.2391 - accuracy:
 0.9130
 Epoch 67/80
 1/1 [=====] - 0s 997us/step - loss: 0.2550 - accuracy:
 0.9275
 Epoch 68/80
 1/1 [=====] - 0s 997us/step - loss: 0.3641 - accuracy:
 0.8841
 Epoch 69/80
 1/1 [=====] - 0s 997us/step - loss: 0.2980 - accuracy:
 0.8551
 Epoch 70/80

```

1/1 [=====] - 0s 997us/step - loss: 0.2459 - accuracy:
0.9275
Epoch 71/80
1/1 [=====] - 0s 998us/step - loss: 0.3395 - accuracy:
0.9130
Epoch 72/80
1/1 [=====] - 0s 997us/step - loss: 0.2106 - accuracy:
0.9565
Epoch 73/80
1/1 [=====] - 0s 0s/step - loss: 0.2134 - accuracy:
0.9130
Epoch 74/80
1/1 [=====] - 0s 998us/step - loss: 0.2345 - accuracy:
0.9275
Epoch 75/80
1/1 [=====] - 0s 997us/step - loss: 0.2167 - accuracy:
0.9130
Epoch 76/80
1/1 [=====] - 0s 1000us/step - loss: 0.1615 - accuracy:
0.9710
Epoch 77/80
1/1 [=====] - 0s 0s/step - loss: 0.1371 - accuracy:
0.9275
Epoch 78/80
1/1 [=====] - 0s 997us/step - loss: 0.2263 - accuracy:
0.9130
Epoch 79/80
1/1 [=====] - 0s 996us/step - loss: 0.1862 - accuracy:
0.9565
Epoch 80/80
1/1 [=====] - 0s 998us/step - loss: 0.0723 - accuracy:
1.0000

```

```

[20]: yr6 = nn_model6.predict(X_tests)

yr6 = le.inverse_transform(yr6)

```

WARNING:tensorflow:5 out of the last 13 calls to <function Model.make_predict_function.<locals>.predict_function at 0x000001FEDB877DC0> triggered tf.function retracing. Tracing is expensive and the excessive number of tracings could be due to (1) creating @tf.function repeatedly in a loop, (2) passing tensors with different shapes, (3) passing Python objects instead of tensors. For (1), please define your @tf.function outside of the loop. For (2), @tf.function has experimental_relax_shapes=True option that relaxes argument shapes that can avoid unnecessary retracing. For (3), please refer to https://www.tensorflow.org/tutorials/customization/performance#python_or_tensor_args and https://www.tensorflow.org/api_docs/python/tf/function for more details.

On va rajouter la colonne “label” au jeu de tests pour comparer avec les résultats théoriques :

```
[22]: yr_label = list()
      for i in range(len(yr)):
          yr_label.append(yr[i])
      df_label = pd.DataFrame(yr_label)
      df_label.columns = ['label']

      df_result = pd.concat([df_emails, df_label['label']], axis=1)
```

```
[23]: df_emails
```

```
[23]:                                     header \
0      Probl me connexion Prod et pr prod postgresql
1      TR: package open source traitement des emails ...
2      RE7_Coverlife/TALEND_BI_COVERLIFE/SID_CVLF_CHA...
3                                     DWH ADP OK - GENERATION_PRD
4      RE7_Coverlife/TALEND_BI_COVERLIFE/SID_CVLF_CHA...
..                                     ...
64      TR: Gestion de la BAL Verlingue Connect
65                                     DWH ADP OK - VERLINGUE_PRD
66                                     DWH ADP OK - GENERATION_PRD
67      [MEP de Version MAJEURE 2021.V2.0] - [FIN DE R...
68      RE7_Coverlife/TALEND_BI_COVERLIFE/SID_CVLF_CHA...

                                     body date \
0      Bonjour,\n\n \n\nLa production et la pr produc... NaT
1      \n\n \n\nDe : Marie Leduc <mleduc@quantmetry... NaT
2      Le traitement SID_CVLF_CHARGEMENT ne s'est pas... NaT
3      Bonjour,\n\nLa mise disposition des donn es ... NaT
4      Le traitement SID_CVLF_CHARGEMENT ne s'est pas... NaT
..                                     ... ...
64      \n\n \n\n-----Rendez-vous d'origine-----\nDe ... NaT
65      Bonjour,\n\nLa mise disposition des donn es ... NaT
66      Bonjour,\n\nLa mise disposition des donn es ... NaT
67      Information de fin de RECETTE post-MEP\n\nnde l... NaT
68      Le traitement SID_CVLF_CHARGEMENT ne s'est pas... NaT

                                     from \
0      Le Moal Mikael
1      L'hotellier Jean-Marie
2      DSI-suivi-production@netdmx.com
3      no-reply@verlingue.fr
4      DSI-suivi-production@netdmx.com
..                                     ...
64      L'hotellier Jean-Marie
65      no-reply@verlingue.fr
```

66 no-reply@verlingue.fr
 67 Fralin Thomas
 68 DSI-suivi-production@netdmx.com

	to	Cc	Cci	\
0	DSI Infra DBA	DSI Data Factory	NaN	
1	Allain Marc;Fresnel Julien		NaN	NaN
2	DSI Data Factory		NaN	NaN
3	DSI Data Factory;Generation Statistiques;Suppo...		NaN	NaN
4	DSI Data Factory		NaN	NaN
..
64	Fresnel Julien		NaN	NaN
65	DSI Data Factory;jvieaud@asi.fr		NaN	NaN
66	DSI Data Factory;Generation Statistiques;Suppo...		NaN	NaN
67	MOA Verlingue;webmoa;DSSI Verlingue;DSI Presta...	DSI Management	NaN	
68	DSI Data Factory		NaN	NaN

	Cat gories	Crit re de diffusion	Importance	...	\
0	NaN	Normal(e)	Normal(e)	...	
1	NaN	Normal(e)	Normal(e)	...	
2	NaN	Normal(e)	Normal(e)	...	
3	NaN	Normal(e)	Normal(e)	...	
4	NaN	Normal(e)	Normal(e)	...	
..	
64	NaN	Normal(e)	Normal(e)	...	
65	NaN	Normal(e)	Normal(e)	...	
66	NaN	Normal(e)	Normal(e)	...	
67	NaN	Normal(e)	Normal(e)	...	
68	NaN	Normal(e)	Normal(e)	...	

	structured_body	\
0	[{'meta': {'date': None, 'from': None, 'to': N...	
1	[{'meta': {'date': None, 'from': 'Marie Leduc ...	
2	[{'meta': {'date': None, 'from': None, 'to': N...	
3	[{'meta': {'date': None, 'from': None, 'to': N...	
4	[{'meta': {'date': None, 'from': None, 'to': N...	
..	...	
64	[{'meta': {'date': None, 'from': None, 'to': N...	
65	[{'meta': {'date': None, 'from': None, 'to': N...	
66	[{'meta': {'date': None, 'from': None, 'to': N...	
67	[{'meta': {'date': None, 'from': None, 'to': N...	
68	[{'meta': {'date': None, 'from': None, 'to': N...	

	last_body	\
0	La production et la pr production PostgreSQL ...	
1	Envoy : mardi 8 juin 2021 08:37 : L'hotelli...	
2	Le traitement SID_CVLF_CHARGEMENT ne s'est pas...	

```

3  La mise disposition des donnees du jour dans...
4  Le traitement SID_CVLF_CHARGEMENT ne s'est pas...
..
64  -----Rendez-vous d'origine-----
65  La mise disposition des donnees du jour dans...
66  La mise disposition des donnees du jour dans...
67  Information de fin de RECETTE post-MEP de la V...
68  Le traitement SID_CVLF_CHARGEMENT ne s'est pas...

```

```

                                clean_body \
0  la production et la prproduction postgresql se...
1  envoy : flag_date_ flag_time_ : l'hotellie...
2  le traitement sid_cvlf_chargement ne s'est pas...
3  la mise disposition des donnees du jour dans le...
4  le traitement sid_cvlf_chargement ne s'est pas...
..
64  rendez-vous d'origine
65  la mise disposition des donnees du jour dans le...
66  la mise disposition des donnees du jour dans le...
67  information de fin de recette post-mep de la v...
68  le traitement sid_cvlf_chargement ne s'est pas...

```

```

                                tokens \
0  [production, prproduction, postgresql, semble,...
1  [envoy, flag_date_, flag_time_, hotellier, fla...
2  [traitement, sid_cvlf_chargement, termin, corr...
3  [mise, disposition, donnees, jour, datawarehous...
4  [traitement, sid_cvlf_chargement, termin, corr...
..
64  [rendez-vous, origine]
65  [mise, disposition, donnees, jour, datawarehous...
66  [mise, disposition, donnees, jour, datawarehous...
67  [information, fin, recette, post-mep, version,...
68  [traitement, sid_cvlf_chargement, termin, corr...

```

```

                                keywords extension hour min \
0  [prproduction, flag_date_, vers, flag_time_, c...      0  NaN NaN
1  [flag_date_, flag_time_, hotellier, flag_name_...      0  NaN NaN
2  [traitement, sid_cvlf_chargement, suivi, produ...      1  NaN NaN
3  [atlas, nombre, flag_cp_, flag_date_, pouvez]          2  NaN NaN
4  [traitement, sid_cvlf_chargement, suivi, produ...      1  NaN NaN
..
64  [origine]                                               0  NaN NaN
65  [nombre, flag_cp_, flag_date_, gnration, pouvez]      2  NaN NaN
66  [nombre, flag_phone_, flag_cp_, flag_date_, po...      2  NaN NaN
67  [flag_time_, tnr, flag_name_, iard, flag_url_]         0  NaN NaN
68  [traitement, sid_cvlf_chargement, suivi, produ...      1  NaN NaN

```

	dayofweek	attachment_type
0	NaN	[0]
1	NaN	[0]
2	NaN	[0]
3	NaN	[0]
4	NaN	[0]
..
64	NaN	[0]
65	NaN	[0]
66	NaN	[0]
67	NaN	[0]
68	NaN	[0]

[69 rows x 26 columns]

[25]: df_result

```
[25]:                                     header \
0      Problème connexion Prod et pr prod postgresql
1      TR: package open source traitement des emails ...
2      RE7_Coverlife/TALEND_BI_COVERLIFE/SID_CVLF_CHA...
3      DWH ADP OK - GENERATION_PRD
4      RE7_Coverlife/TALEND_BI_COVERLIFE/SID_CVLF_CHA...
..      ...
64      TR: Gestion de la BAL Verlingue Connect
65      DWH ADP OK - VERLINGUE_PRD
66      DWH ADP OK - GENERATION_PRD
67      [MEP de Version MAJEURE 2021.V2.0] - [FIN DE R...
68      RE7_Coverlife/TALEND_BI_COVERLIFE/SID_CVLF_CHA...

                                     body date \
0      Bonjour,\n\n \n\nLa production et la pr produc... NaT
1      \n\n \n\nDe : Marie Leduc <mleduc@quantmetry... NaT
2      Le traitement SID_CVLF_CHARGEMENT ne s'est pas... NaT
3      Bonjour,\n\nLa mise disposition des donn es ... NaT
4      Le traitement SID_CVLF_CHARGEMENT ne s'est pas... NaT
..      ...
64      \n\n \n\n-----Rendez-vous d'origine-----\nDe ... NaT
65      Bonjour,\n\nLa mise disposition des donn es ... NaT
66      Bonjour,\n\nLa mise disposition des donn es ... NaT
67      Information de fin de RECETTE post-MEP\n\nnde l... NaT
68      Le traitement SID_CVLF_CHARGEMENT ne s'est pas... NaT

                                     from \
0      Le Moal Mikael
1      L'hotellier Jean-Marie
```

```

2 DSI-suivi-production@netdmx.com
3 no-reply@verlingue.fr
4 DSI-suivi-production@netdmx.com
..
64 L'hotellier Jean-Marie
65 no-reply@verlingue.fr
66 no-reply@verlingue.fr
67 Fralin Thomas
68 DSI-suivi-production@netdmx.com

```

```

                                to Cc Cci \
0 DSI Infra DBA DSI Data Factory NaN
1 Allain Marc;Fresnel Julien NaN NaN
2 DSI Data Factory NaN NaN
3 DSI Data Factory;Generation Statistiques;Suppo... NaN NaN
4 DSI Data Factory NaN NaN
..
64 Fresnel Julien NaN NaN
65 DSI Data Factory;jvieaud@asi.fr NaN NaN
66 DSI Data Factory;Generation Statistiques;Suppo... NaN NaN
67 MOA Verlingue;webmoa;DSSI Verlingue;DSI Presta... DSI Management NaN
68 DSI Data Factory NaN NaN

```

```

Cat gories Crit re de diffusion Importance ... \
0 NaN Normal(e) Normal(e) ...
1 NaN Normal(e) Normal(e) ...
2 NaN Normal(e) Normal(e) ...
3 NaN Normal(e) Normal(e) ...
4 NaN Normal(e) Normal(e) ...
..
64 NaN Normal(e) Normal(e) ...
65 NaN Normal(e) Normal(e) ...
66 NaN Normal(e) Normal(e) ...
67 NaN Normal(e) Normal(e) ...
68 NaN Normal(e) Normal(e) ...

```

```

                                last_body \
0 La production et la pr production PostgreSQL ...
1 Envoy : mardi 8 juin 2021 08:37 : L'hotelli...
2 Le traitement SID_CVLF_CHARGEMENT ne s'est pas...
3 La mise disposition des donn es du jour dans...
4 Le traitement SID_CVLF_CHARGEMENT ne s'est pas...
..
64 -----Rendez-vous d'origine-----
65 La mise disposition des donn es du jour dans...
66 La mise disposition des donn es du jour dans...
67 Information de fin de RECETTE post-MEP de la V...

```

68 Le traitement SID_CVLF_CHARGEMENT ne s'est pas...

```

                                clean_body \
0  la production et la prproduction postgresql se...
1  envoy : flag_date_  flag_time_ : l'hotellie...
2  le traitement sid_cvlf_chargement ne s'est pas...
3  la mise disposition des donnees du jour dans le...
4  le traitement sid_cvlf_chargement ne s'est pas...
..
64                                rendez-vous d'origine
65  la mise disposition des donnees du jour dans le...
66  la mise disposition des donnees du jour dans le...
67  information de fin de recette post-mep de la v...
68  le traitement sid_cvlf_chargement ne s'est pas...

```

```

                                tokens \
0  [production, prproduction, postgresql, semble,...
1  [envoy, flag_date_, flag_time_, hotellier, fla...
2  [traitement, sid_cvlf_chargement, termin, corr...
3  [mise, disposition, donnees, jour, datawarehous...
4  [traitement, sid_cvlf_chargement, termin, corr...
..
64                                [rendez-vous, origine]
65  [mise, disposition, donnees, jour, datawarehous...
66  [mise, disposition, donnees, jour, datawarehous...
67  [information, fin, recette, post-mep, version,...
68  [traitement, sid_cvlf_chargement, termin, corr...

```

```

                                keywords  extension hour min \
0  [prproduction, flag_date_, vers, flag_time_, c...      0  NaN NaN
1  [flag_date_, flag_time_, hotellier, flag_name_...      0  NaN NaN
2  [traitement, sid_cvlf_chargement, suivi, produ...      1  NaN NaN
3    [atlas, nombre, flag_cp_, flag_date_, pouvez]        2  NaN NaN
4  [traitement, sid_cvlf_chargement, suivi, produ...      1  NaN NaN
..
64                                [origine]                0  NaN NaN
65  [nombre, flag_cp_, flag_date_, gnration, pouvez]      2  NaN NaN
66  [nombre, flag_phone_, flag_cp_, flag_date_, po...      2  NaN NaN
67    [flag_time_, tnr, flag_name_, iard, flag_url_]       0  NaN NaN
68  [traitement, sid_cvlf_chargement, suivi, produ...      1  NaN NaN

```

	dayofweek	attachment_type	label
0	NaN	[0]	production
1	NaN	[0]	production
2	NaN	[0]	production
3	NaN	[0]	datawarehouse
4	NaN	[0]	production


```

..      ...      ...      ...
64      NaN      [0]      production
65      NaN      [0]      datawarehouse
66      NaN      [0]      datawarehouse
67      NaN      [0]      production
68      NaN      [0]      production

```

[69 rows x 27 columns]

Les résultats sont curieusement moins exacts sur les **Notebooks** par rapport aux tests effectués sur via **PyCharm**

Résultats obtenus via Pycharm

```
[26]: df_emails_pycharm = pd.read_csv(r"C:\Users\JUFL\Desktop\test3.CSV", delimiter=";"
    ↪)
```

```
[28]: df_emails_pycharm
```

```

[28]:      Unnamed: 0      header \
0      0      Problme connexion Prod et pr prod postgresql
1      1      TR: package open source traitement des emails ...
2      2      RE7_Coverlife/TALEND_BI_COVERLIFE/SID_CVLF_CHA...
3      3      DWH ADP OK - GENERATION_PRD
4      4      RE7_Coverlife/TALEND_BI_COVERLIFE/SID_CVLF_CHA...
..      ...      ...
64     64      TR: Gestion de la BAL Verlingue Connect
65     65      DWH ADP OK - VERLINGUE_PRD
66     66      DWH ADP OK - GENERATION_PRD
67     67      [MEP de Version MAJEURE 2021.V2.0] - [FIN DE R...
68     68      RE7_Coverlife/TALEND_BI_COVERLIFE/SID_CVLF_CHA...

      body  date \
0  Bonjour,\n\n \n\nLa production et la pr produc...  NaN
1  \n\n \n\nDe : Marie Leduc <mleduc@quantmetry...  NaN
2  Le traitement SID_CVLF_CHARGEMENT ne s'est pas...  NaN
3  Bonjour,\n\nLa mise disposition des donn es ...  NaN
4  Le traitement SID_CVLF_CHARGEMENT ne s'est pas...  NaN
..      ...      ...
64  \n\n \n\n-----Rendez-vous d'origine-----\nDe ...  NaN
65  Bonjour,\n\nLa mise disposition des donn es ...  NaN
66  Bonjour,\n\nLa mise disposition des donn es ...  NaN
67  Information de fin de RECETTE post-MEP\n\nnde l...  NaN
68  Le traitement SID_CVLF_CHARGEMENT ne s'est pas...  NaN

      from \
0      Le Moal Mikael
1      L'hotellier Jean-Marie

```

```

2 DSI-suivi-production@netdmx.com
3 no-reply@verlingue.fr
4 DSI-suivi-production@netdmx.com
..
64 L'hotellier Jean-Marie
65 no-reply@verlingue.fr
66 no-reply@verlingue.fr
67 Fralin Thomas
68 DSI-suivi-production@netdmx.com

```

```

                                to Cc Cci \
0 DSI Infra DBA DSI Data Factory NaN
1 Allain Marc;Fresnel Julien NaN NaN
2 DSI Data Factory NaN NaN
3 DSI Data Factory;Generation Statistiques;Suppo... NaN NaN
4 DSI Data Factory NaN NaN
..
64 Fresnel Julien NaN NaN
65 DSI Data Factory;jvieaud@asi.fr NaN NaN
66 DSI Data Factory;Generation Statistiques;Suppo... NaN NaN
67 MOA Verlingue;webmoa;DSSI Verlingue;DSI Presta... DSI Management NaN
68 DSI Data Factory NaN NaN

```

```

Cat gories Crit re de diffusion ... \
0 NaN Normal(e) ...
1 NaN Normal(e) ...
2 NaN Normal(e) ...
3 NaN Normal(e) ...
4 NaN Normal(e) ...
..
64 NaN Normal(e) ...
65 NaN Normal(e) ...
66 NaN Normal(e) ...
67 NaN Normal(e) ...
68 NaN Normal(e) ...

```

```

                                structured_body \
0 [{'meta': {'date': None, 'from': None, 'to': N...
1 [{'meta': {'date': None, 'from': 'Marie Leduc ...
2 [{'meta': {'date': None, 'from': None, 'to': N...
3 [{'meta': {'date': None, 'from': None, 'to': N...
4 [{'meta': {'date': None, 'from': None, 'to': N...
..
64 [{'meta': {'date': None, 'from': None, 'to': N...
65 [{'meta': {'date': None, 'from': None, 'to': N...
66 [{'meta': {'date': None, 'from': None, 'to': N...
67 [{'meta': {'date': None, 'from': None, 'to': N...

```

```

68 [{"meta": {"date": None, "from": None, "to": N...

                                last_body \
0   La production et la pr production PostgreSQL ...
1   Envoy : mardi 8 juin 2021 08:37 : L'hotelli...
2   Le traitement SID_CVLF_CHARGEMENT ne s'est pas...
3   La mise disposition des donnees du jour dans...
4   Le traitement SID_CVLF_CHARGEMENT ne s'est pas...
..
64   -----Rendez-vous d'origine-----
65   La mise disposition des donnees du jour dans...
66   La mise disposition des donnees du jour dans...
67   Information de fin de RECETTE post-MEP de la V...
68   Le traitement SID_CVLF_CHARGEMENT ne s'est pas...

                                clean_body \
0   la production et la prproduction postgresql se...
1   envoy : flag_date_ flag_time_ : l'hotellie...
2   le traitement sid_cvlf_chargement ne s'est pas...
3   la mise disposition des donnees du jour dans le...
4   le traitement sid_cvlf_chargement ne s'est pas...
..
64   rendez-vous d'origine
65   la mise disposition des donnees du jour dans le...
66   la mise disposition des donnees du jour dans le...
67   information de fin de recette post-mep de la v...
68   le traitement sid_cvlf_chargement ne s'est pas...

                                tokens extension hour min \
0   ['production', 'prproduction', 'postgresql', '...' 0 NaN NaN
1   ['envoy', 'flag_date_', 'flag_time_', 'hotelli...' 0 NaN NaN
2   ['traitement', 'sid_cvlf_chargement', 'termin'...' 1 NaN NaN
3   ['mise', 'disposition', 'donnees', 'jour', 'dat...' 2 NaN NaN
4   ['traitement', 'sid_cvlf_chargement', 'termin'...' 1 NaN NaN
..
64   ['rendez-vous', 'origine'] 0 NaN NaN
65   ['mise', 'disposition', 'donnees', 'jour', 'dat...' 2 NaN NaN
66   ['mise', 'disposition', 'donnees', 'jour', 'dat...' 2 NaN NaN
67   ['information', 'fin', 'recette', 'post-mep', '...' 0 NaN NaN
68   ['traitement', 'sid_cvlf_chargement', 'termin'...' 1 NaN NaN

dayofweek attachment_type label
0   NaN [0] production
1   NaN [0] personnel
2   NaN [0] traitement
3   NaN [0] datawarehouse
4   NaN [0] traitement

```

```

..      ...      ...      ...
64      NaN      [0]      rdv
65      NaN      [0]      datawarehouse
66      NaN      [0]      datawarehouse
67      NaN      [0]      production
68      NaN      [0]      traitement

```

[69 rows x 26 columns]

Comparons les résultats obtenus via Pycharm avec les résultats théoriques:

```

[31]: df_emails_pycharm.loc[df_emails_pycharm['label']==df_emails['label'], 'Match']=1
      ↪1
      df_emails_pycharm.loc[df_emails_pycharm['label']!=df_emails['label'], 'Match']=0
      ↪0

```

```

[35]: N = len(df_emails_pycharm['Match'])

```

```

[39]: S = df_emails_pycharm['Match'].sum()

```

```

[42]: percent = (S/N)*100

      print(str(round(percent,2)) + ' % of accuracy')

```

89.86 % of accuracy