


Réalisez un dashboard et  
assurez une veille technique



# Sommaire

- ❖ Problématique
  - ❖ API
  - ❖ Dashboard
  - ❖ Modélisation pour la veille Technologique
  - ❖ Conclusion et Suite du Projet
- 

# Problématique : Contexte



Société financière d'offre de crédit à la consommation pour la clientèle ayant peu ou pas d'historique de prêt.

## Mission :

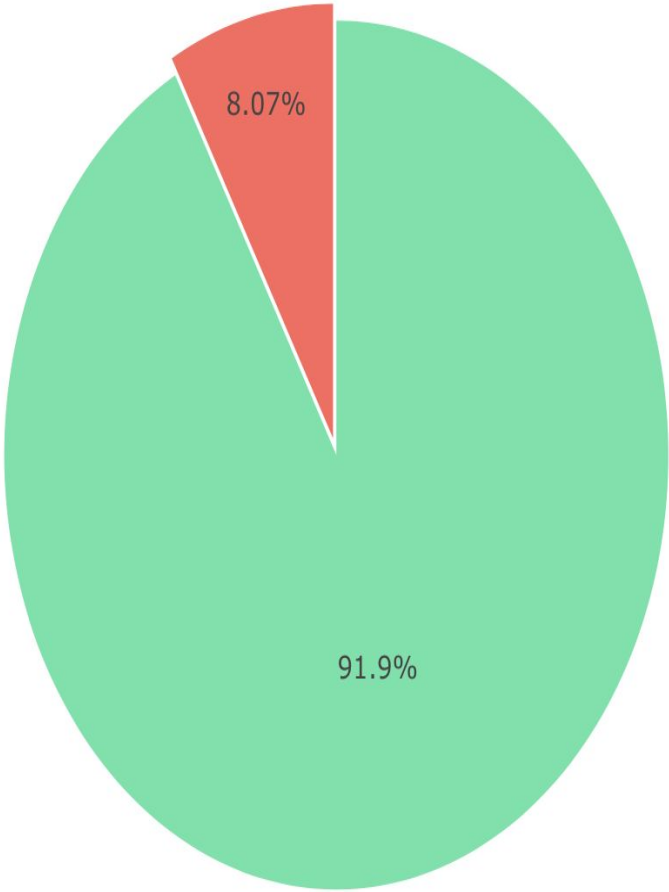
- ❖ Une fois que la mise en place de l'API fut un succès. La suite consiste à créer un Dashboard interactif afin de Montrer au client les informations le concernant grâce à l'interactivité.
- ❖ Au vu de l'avancé du projet, réaliser une veille technologique sur des problématiques de données texte (NLP) ou de données d'images.

## Objectif :

- ❖ Étayer la décision d'accorder ou non un prêt.
- ❖ Améliorer la relation avec le client en faisant preuve de transparence.
- ❖ Mettre en place un état de l'art sur une technique datant de moins de 5 ans, présentée dans un article.



# Dashboard Modélisation : Variable Cible



■ Non-Defaulter  
■ Defaulter

Variable Cible Binaire	
<b>Classe 0</b> Clients <b>non défaillants</b> 92% majoritaire	<b>Classe 1</b> Clients <b>défaillants</b> 8% minoritaire



Classification Binaire avec  
Classe déséquilibrée

# Dashboard Modélisation : Choix des Métriques

## Rappel:

Classe 0 négative = non défaillant

Classe 1 positive = défaillant

## Matrice de Confusion

Classe réelle	+	<b>TP</b> Vrais positifs	<b>FN</b> Faux négatifs
	-	<b>FP</b> Faux positifs	<b>TN</b> Vrais négatifs
		+	-
		Classe prédite	

## Minimiser les pertes argents :



**minimiser** le nombre de **faux positifs**  
**maximiser** la métrique **Précision**



**minimiser** le nombre de **faux négatifs**  
**maximiser** les métriques **Recall** ou **Fbeta**

# Dashboard Modélisation : Tracking avec MLFLOW

mlflow2.11.1

ExperimentsModels

Experiments

Search Experiments

Credit\_Score

metrics.rmse < 1 and params.model = "tree"

Time created

State: Active

Datasets

Sort: Created

Columns

Group by

Table

Chart

Evaluation

Experimental

	Run Name	Created	Dataset	Duration	Source	Models
	dummy_hyperparam_base	2 hours ago	-	3.7s	ipykerne...	-
	lgbm_hyperparam_base	2 hours ago	-	8.8min	ipykerne...	-
	lgbm_hyperparam_base_bal	2 hours ago	-	9.7min	ipykerne...	-
	xgb_hyperparam_base	2 hours ago	-	10.2min	ipykerne...	-
	xgb_hyperparam_base_bal	2 hours ago	-	10.1min	ipykerne...	-
	lr_hyperparam_base_bal	2 hours ago	-	15.5min	ipykerne...	-
	lgbm_optuna_opt_1	1 hour ago	-	8.2min	ipykerne...	-
	lgbm_optuna_opt_2	1 hour ago	-	4.9min	ipykerne...	-
	lgbm_optuna_smote_opt_3	1 hour ago	-	14.9min	ipykerne...	-
	lgbm_optuna_smote_opt_4_train	1 hour ago	-	7.0min	ipykerne...	-
	lgbm_optuna_smote_opt_4_train	56 minutes ago	-	7.2min	ipykerne...	-
	lgbm_optuna_smote_opt_4	49 minutes ago	-	13.5min	ipykerne...	-

Credit\_Score

Provide Feedback

Add Description

Share

metrics.rmse < 1 and params.model = "tree"

Time created

State: Active

Datasets

Sort: Created

Group by

Table

Chart

Evaluation

Experimental

Run Name

dummy\_hyperparam\_b...

lgbm\_hyperparam\_base

lgbm\_hyperparam\_base

xgb\_hyperparam\_base

xgb\_hyperparam\_base\_bal

lr\_hyperparam\_base\_bal

lgbm\_optuna\_opt\_1

lgbm\_optuna\_opt\_2

lgbm\_optuna\_smote\_o...

lgbm\_optuna\_smote\_o...

lgbm\_optuna\_smote\_o...

lgbm\_optuna\_smote\_o...

metier\_score

Comparing first 10 runs

pc\_auc

Comparing first 10 runs

precision

Comparing first 10 runs

recall

Comparing first 10 runs

12 matching runs

Credit\_Score

lgbm\_optuna\_opt\_2

Overview

Model metrics

System metrics

Artifacts

Status

Run ID

Duration

Datasets used

Tags

Source

Logged models

Registered models

Parameters (2)

Search parameters

Parameter

Value

model\_params

model\_type

Metrics (7)

Search metrics

Metric

Value

execution\_time

precision

f1\_score

roc\_auc

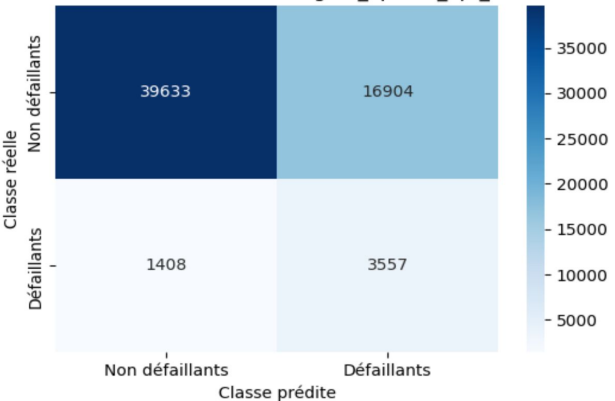
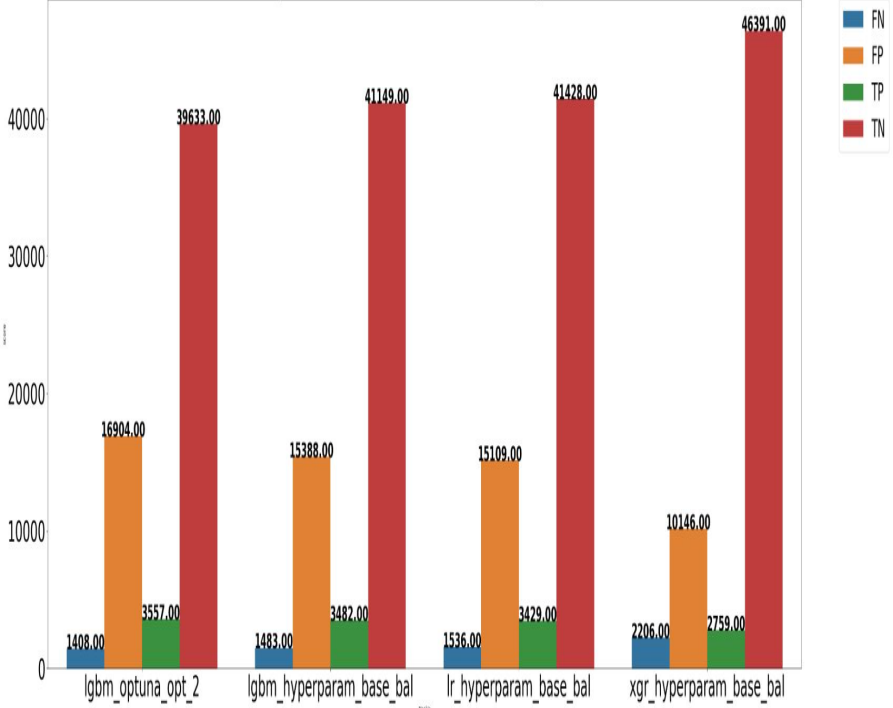
metier\_score

recall

pc\_auc

# Dashboard Modélisation : Bilan

Comparaison des scores des 4 meilleurs algorithmes



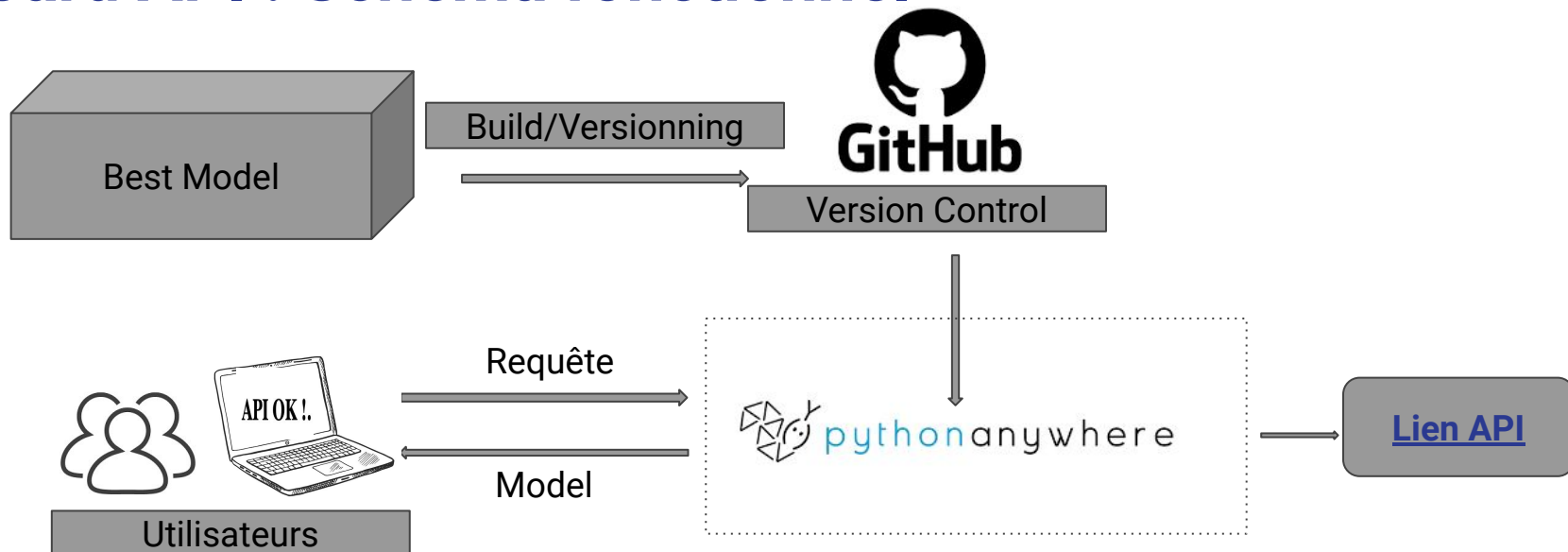
Modèle	FN	FP	TP	TN
lgbm_optuna_opt_2	1408	16904	3557	39633
lgbm_hyperparam_base_bal	1483	15388	3482	41149
lr_hyperparam_base_bal	1536	15109	3429	41428
xgr_hyperparam_base_bal	2206	10146	2759	46391
lgbm_optuna_smote_opt_4_train	4490	950	475	55587
dummy_hyperparam_base	4570	4587	395	51950
dummy_hyperparam_base	4570	4587	395	51950
lgbm_optuna_smote_opt_4	4588	722	51949	55815
lgbm_optuna_smote_opt_3	4781	195	51756	56342
lgbm_hyperparam_base	4783	133	182	56404

RETENU pour API



lgbm\_otuna\_opt\_2

# Dashboard API : Schéma fonctionnel



Non sécurisé — bbbkalilunix.pythonanywhere.com

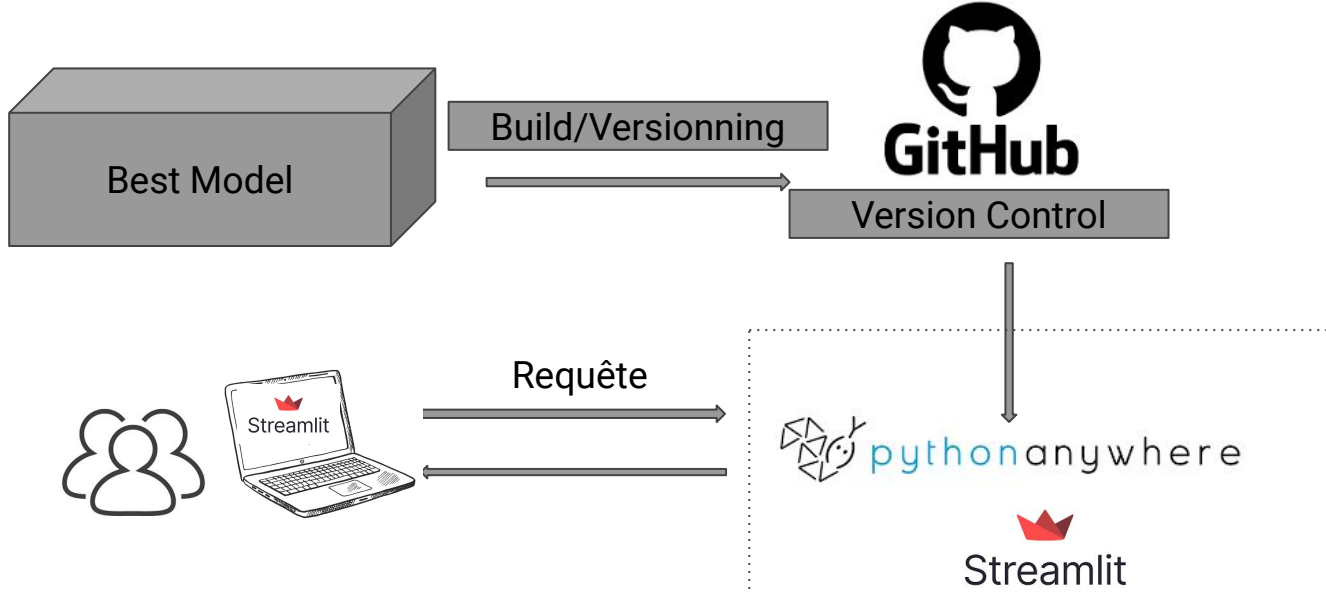


## Implémenter Un Modèle de Scoring :

API OK !.



# Dashboard



Definir ID de client:

ID de client

101099

☐ Informations générales

☒ Analyse de la demande de crédit

☐ Analyse des features de client

Informations générales

Revenu moyen (USD):

182881.12

Montant moyen du prêt (USD):

523164.7

## Prêt à dépenser

### DASHBOARD

Informations sur les clients

Vous pouvez ajouter ou enlever une donnée présente dans cette liste:

TARGET\_NEIGH... x

NAME\_INCOME\_... x

PREV\_APP\_CHA... x

OCCUPATION\_TY... x

OBS\_30\_CNT\_S... x

PREV\_APP\_NAM... x

PREV\_APP\_COD... x

ORGANIZATION\_... x

OCCUPATION\_TY... x

PREV\_APP\_NAM... x

SK_ID_CURR	TARGET_NEIGHBORS_500_MEAN	NAME_INCOME_TYPE_WORKING	PREV_APP_CHANNEL_TYPE_CHANNEL_OF_CORPORATE_SALES_MEAN
119,115	0.06	0	0
119,232	0.088	0	0
119,391	0.078	1	0
120,759	0.026	0	0
120,921	0.014	0	0
120,936	0.054	1	0

# Dashboard : Accueil

×

Prêt à dépenser

Definir ID de client:

ID de client

100038

▼

☐ Informations générales

☐ Analyse de la demande de crédit

☒ Analyse des features de client

Informations générales

Revenu moyen (USD) :

182881.12

Montant moyen du prêt (USD) :

523164.7

Deploy

⋮

Prêt à dépenser

DASHBOARD

Informations sur les clients

Vous pouvez ajouter ou enlever une donnée présente dans cette liste:

TARGET\_NEIGH... ×

NAME\_INCOME\_... ×

PREV\_APP\_CHA... ×

OCCUPATION\_TY... ×

OBS\_30\_CNT\_S... ×

PREV\_APP\_NAM... ×

PREV\_APP\_COD... ×

ORGANIZATION\_... ×

OCCUPATION\_TY... ×

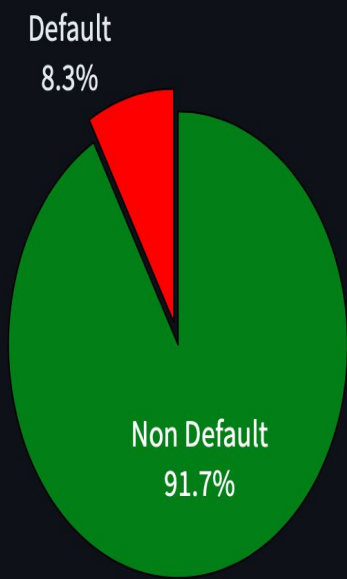
PREV\_APP\_NAM... ×

⊗ ▼

SK_ID_CURR	TARGET_NEIGHBORS_500_MEAN	NAME_INCOME_TYPE_WORKING	PREV_APP_CHANNEL_TYPE_CHANNEL_OF_CORPORATE_SALES_MEAN
100,038	0.122	1	0
101,099	0.06	0	0
102,158	0.188	0	0
102,273	0.126	0	0
103,184	0.034	0	0
103,258	0.046	1	0
103,286	0.054	0	0
103,336	0.048	0	0
103,484	0.022	1	0
104,391	0.024	0	0

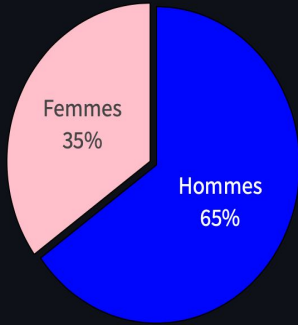
# Dashboard : Informations générales sur les clients

Proportion de 'TARGET'



■ Non Defe  
■ Default


Proportion de 'GENDRE'



■ Hommes  
■ Femmes



# Dashboard : Informations générales sur un client



Definir ID de client:

ID de client

100038

☒ Informations générales

☐ Analyse de la demande de crédit

☐ Analyse des features de client

Informations générales

Revenu moyen (USD):

182881.12

Montant moyen du prêt (USD):

523164.7

Données du client, demande 100038

Vous pouvez afficher 5 données maximum parmi cette liste:

NAME\_INCOME\_...

TARGET\_NEIGH...

FLAG\_DOCUMENT...

FLOORSMAX\_AVG

FLOORSMIN\_AVG

PREV\_APP\_NAME\_SELLER\_INDUSTRY\_INDUSTRY\_MEAN

PREV\_APP\_NAME\_CONTRACT\_TYPE\_CASH\_LOANS\_SUM

NAME\_FAMILY\_STATUS\_SINGLE\_NOT\_MARRIED

PREV\_APP\_NAME\_TYPE\_SUITE\_FAMILY\_SUM

PREV\_APP\_NAME\_PRODUCT\_TYPE\_WALK\_IN\_SUM

PREV\_APP\_NAME\_GOODS\_CATEGORY\_DIRECT\_SALES\_MEAN

Valeur normalisée

NAME\_INCOME\_TYPE\_WORKING

TARGET\_NEIGHBORS\_500\_MEAN

FLAG\_DOCUMENT\_3

FLOORSMAX\_AVG

1


0.8

0.6

0.4

0.2

0



Feature	Valeur normalisée
NAME_INCOME_TYPE_WORKING	1.0
TARGET_NEIGHBORS_500_MEAN	0.12
FLAG_DOCUMENT_3	1.0
FLOORSMAX_AVG	0.18

# Dashboard : Demande de prêt pour un client



## Definir ID de client:

ID de client

100038

- ☐ Informations générales
- ☒ Analyse de la demande de crédit
- ☐ Analyse des features de client

## Informations générales

Revenu moyen (USD) :

182881.12

Montant moyen du prêt (USD) :

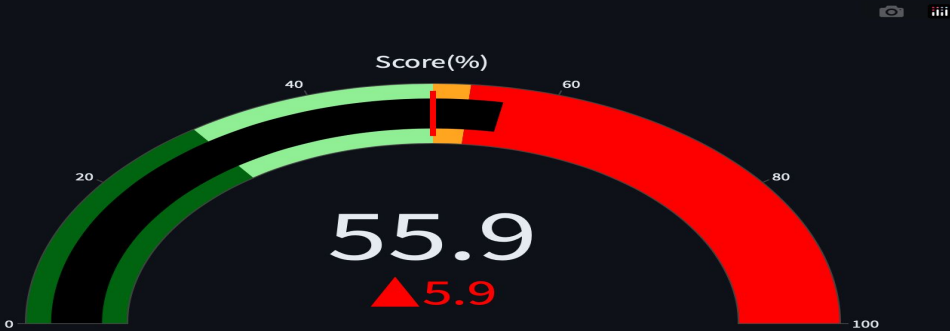
523164.7

## Décision sur la demande de prêt

Demande de prêt ID: 100038

Probabilité de défauts de remboursement: 55.92 %

Demande de prêt refusée!



### Informations complémentaires

Le retour de l'API de prédiction donne un score entre 0 et 100% qui représente la probabilité de refus de prêt. Trois cas de figure sont alors possibles:

1. Le score est en dessous de 50% → la demande de prêt est acceptée.
2. Le score est entre 50 et 53% → la demande de prêt est refusée mais peut être discutée avec le conseiller pour éventuellement l'accepter (grâce notamment à l'onglet 'analyse des features clients').
3. Le score est au dessus de 53% → la demande de prêt est refusée.

# Dashboard : Analyses univariées



## Definir ID de client:

ID de client

101099

- ☐ Informations générales
- ☐ Analyse de la demande de crédit
- ☒ Analyse des features de client

## Informations générales

Revenu moyen (USD):

182881.12

Montant moyen du prêt (USD):

523164.7

### Analyse des features clients

#### Analyse univariée

Sélectionner un graphique

Boxplot

Vous pouvez ajouter ou enlever une donnée présente dans cette liste:

TARGET\_NEIGH... x NAME\_INCOME,... x PREV\_APP\_CHA... x OCCUPATION\_TY... x OBS\_30\_CNT\_S...

value

TARGET\_NEIGHBORS\_500\_MEAN NAME\_INCOME\_TYPE\_WORKING PREV\_APP\_CHANNEL\_TYPE\_CHANNEL\_OF\_CORPORATE\_SALES\_MEAN OCCUPATION\_TYPE\_CORE\_STAFF OBS\_30\_CNT\_SOCIAL\_CIRCLE

Valeur normalisée

Features

Tous les clients ID\_prêt\_client\_selectionné

# Dashboard : Analyses bivariées



## Definir ID de client:

ID de client

101099

- ☐ Informations générales
- ☐ Analyse de la demande de crédit
- ☒ Analyse des features de client

## Informations générales

Revenu moyen (USD) :

182881.12

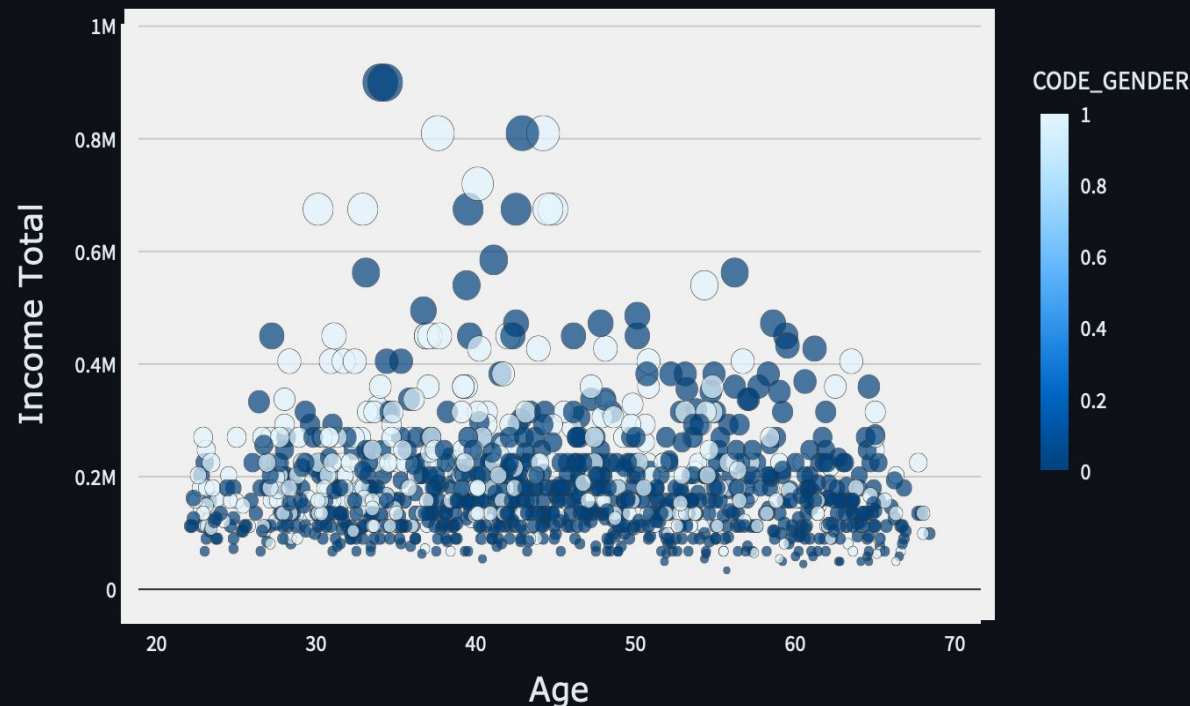
Montant moyen du prêt (USD) :

523164.7



## Analyse bivée

### Relation Âge / Revenu Total





# Dashboard : Features Importances



## Definir ID de client:

ID de client

101099

▼

- ☐ Informations générales
- ☐ Analyse de la demande de crédit
- ☒ Analyse des features de client

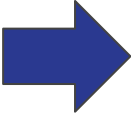
## Informations générales

Revenu moyen (USD):

182881.12

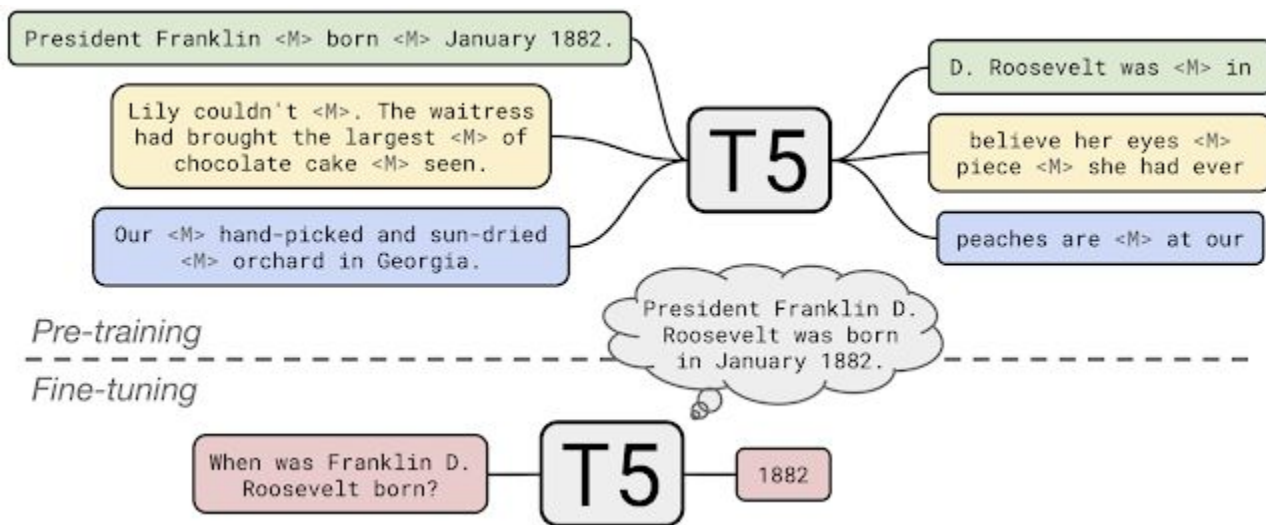
Montant moyen du prêt (USD):

523164.7





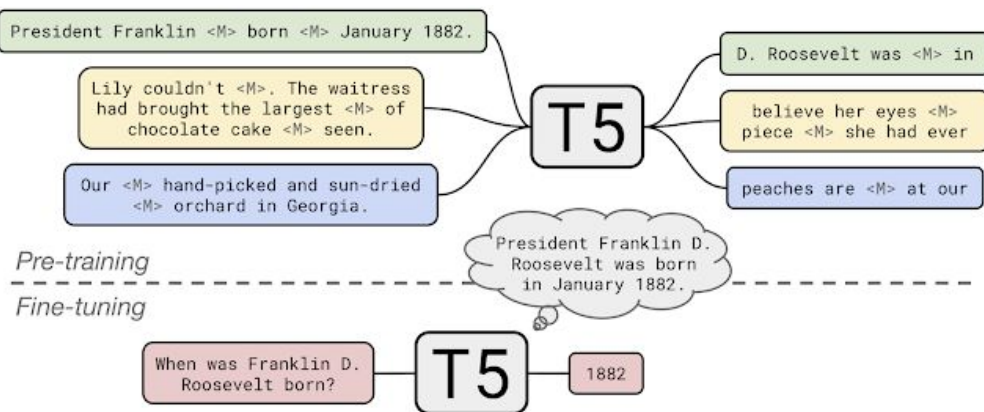
# Veille Technologique: Modele T5



Par exemple, pour une tâche de classification de texte, le texte d'entrée pourrait être une phrase ou un paragraphe, et le texte de sortie serait une étiquette de classe correspondante. De même, pour une tâche de génération de texte, le texte d'entrée pourrait être une phrase ou une description, et le texte de sortie serait la séquence de mots générée.

Le modèle T5, ou "Text-To-Text Transfer Transformer", est une architecture de réseau neuronal développée par Google AI. Elle a été introduite dans un article de recherche intitulé "Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer" publié en 2020 (Raffel et al., 2020).

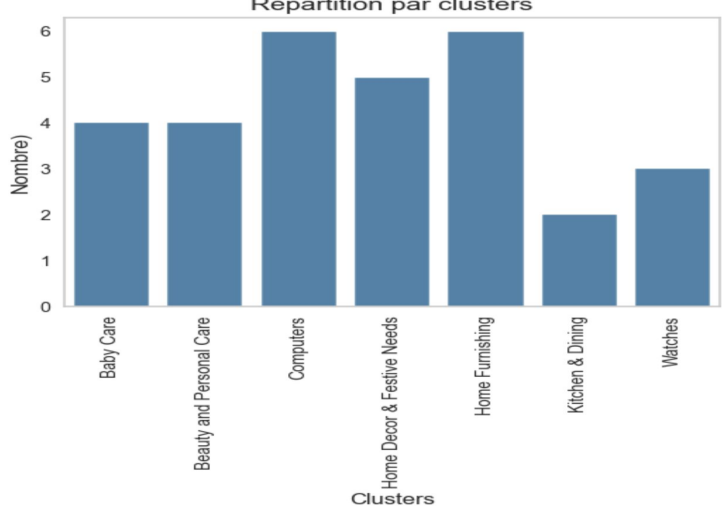
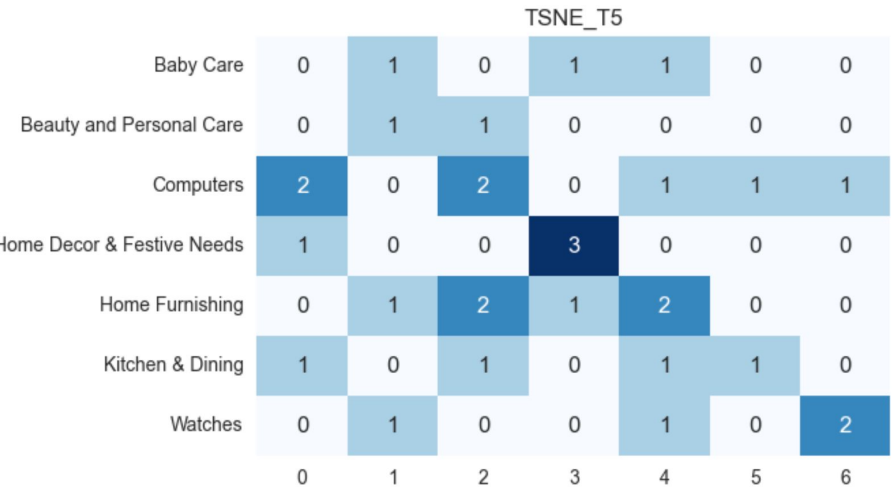
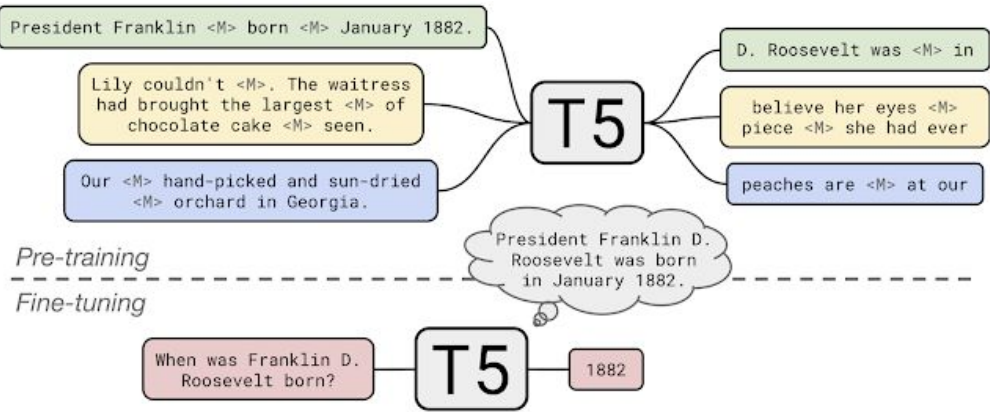
# Modèle T5 : Concept



- ❖ **Un cadre unifié** : formulées sous forme de tâches de «texte vers texte».
- ❖ **Architecture Transformer** : capturer les dépendances à longue portée dans les séquences de texte.

- ❖ **Pré-entraînement et Fine-tuning** : T5 est pré-entraîné sur un large corpus de données textuelles à l'aide d'une tâche d'auto-encodage textuel.
- ❖ **Flexibilité** : Peut être appliqué à une grande variété de tâches de traitement du langage naturel en reformulant simplement la tâche d'entrée et de sortie.
- ❖ **Performances Élevées** : performances impressionnantes sur une gamme de tâches de NLP

# Modèle T5 : Résultats



	Model	Accuracy	Recall	Prec.	F1	Kappa	MCC
lr	Logistic Regression	0.2667	0.2667	0.2619	0.2634	0.2625	0.2667
lda	Linear Discriminant Analysis	0.3429	0.3429	0.2214	0.2595	0.2301	0.2457
rf	Random Forest Classifier	0.3000	0.3000	0.2333	0.2500	0.1857	0.1985
et	Extra Trees Classifier	0.3000	0.3000	0.2405	0.2595	0.1844	0.1937
qda	Quadratic Discriminant Analysis	0.2857	0.2857	0.1667	0.2038	0.1631	0.1769
nb	Naive Bayes	0.2714	0.2714	0.1619	0.1952	0.1491	0.1640
ridge	Ridge Classifier	0.2714	0.2714	0.1405	0.1800	0.1528	0.1717
dt	Decision Tree Classifier	0.2571	0.2571	0.1976	0.2167	0.1273	0.1376
knn	K Neighbors Classifier	0.2429	0.2429	0.1940	0.2033	0.1243	0.1384
gbc	Gradient Boosting Classifier	0.2286	0.2286	0.1560	0.1724	0.0967	0.1052
xgboost	Extreme Gradient Boosting	0.2286	0.2286	0.1714	0.1833	0.1011	0.1077
svm	SVM - Linear Kernel	0.2143	0.2143	0.1024	0.1343	0.0822	0.0943
lightgbm	Light Gradient Boosting Machine	0.1857	0.1857	0.1405	0.1500	0.0459	0.0506
dummy	Dummy Classifier	0.1571	0.1571	0.0265	0.0448	0.0000	0.0000
ada	Ada Boost Classifier	0.1429	0.1429	0.0560	0.0754	-0.0088	-0.0076

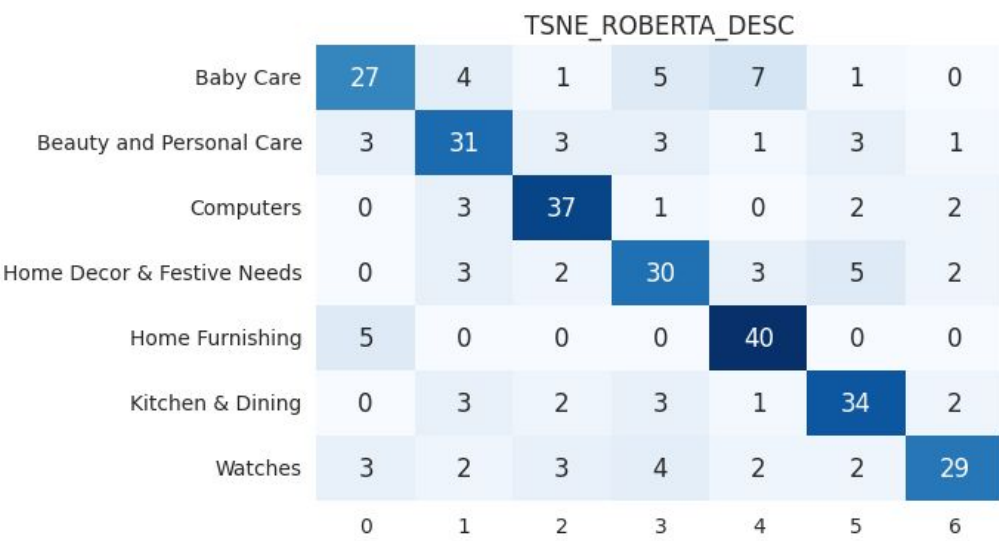
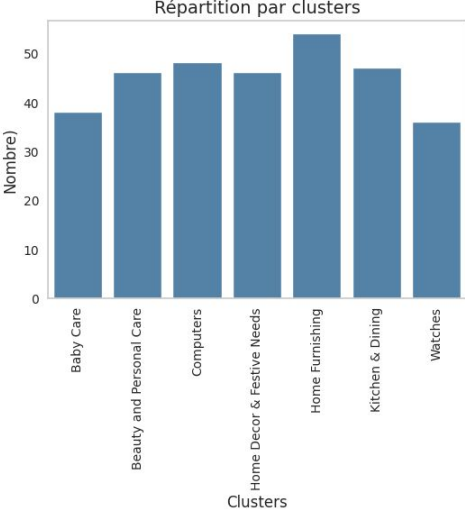
'Précision: 36.67%'

	Model	Accuracy	Recall	Prec.	F1	Kappa	MCC
0	Logistic Regression	0.2667	0.2667	0.2194	0.2402	0.1339	0.1357

# Anciens Modèles : Résultats BERT

	Model	Accuracy	Recall	Prec.	F1	Kappa	MCC
et	Extra Trees Classifier	0.7575	0.7575	0.7409	0.7538	0.8838	0.8853
rf	Random Forest Classifier	0.7198	0.7198	0.7302	0.7167	0.6729	0.6754
lightgbm	Light Gradient Boosting Machine	0.6980	0.6980	0.7055	0.6952	0.6475	0.6496
knn	K Neighbors Classifier	0.6775	0.6775	0.6923	0.6752	0.6237	0.6266
xgboost	Extreme Gradient Boosting	0.6749	0.6749	0.6870	0.6705	0.6205	0.6231
dt	Decision Tree Classifier	0.6680	0.6680	0.6774	0.6629	0.6125	0.6155
gbc	Gradient Boosting Classifier	0.6477	0.6477	0.6557	0.6448	0.5888	0.5909
qda	Quadratic Discriminant Analysis	0.4012	0.4012	0.4033	0.3675	0.3016	0.3109
nb	Naive Bayes	0.3876	0.3876	0.3784	0.3574	0.2856	0.2922
ada	Ada Boost Classifier	0.3511	0.3511	0.3574	0.3368	0.2434	0.2476
lr	Logistic Regression	0.2625	0.2625	0.2555	0.2228	0.1399	0.1471
lda	Linear Discriminant Analysis	0.2585	0.2585	0.2515	0.2183	0.1351	0.1423
ridge	Ridge Classifier	0.2230	0.2230	0.1253	0.1500	0.0937	0.1044
svm	SVM - Linear Kernel	0.2027	0.2027	0.2585	0.1755	0.0698	0.0766
dummy	Dummy Classifier	0.1361	0.1361	0.0185	0.0326	0.0000	0.0000

	Model	Accuracy	Recall	Prec.	F1	Kappa	MCC
0	Extra Trees Classifier	0.7238	0.7238	0.7253	0.7215	0.6778	0.6787



# Anciens Modèles : Résultats USE

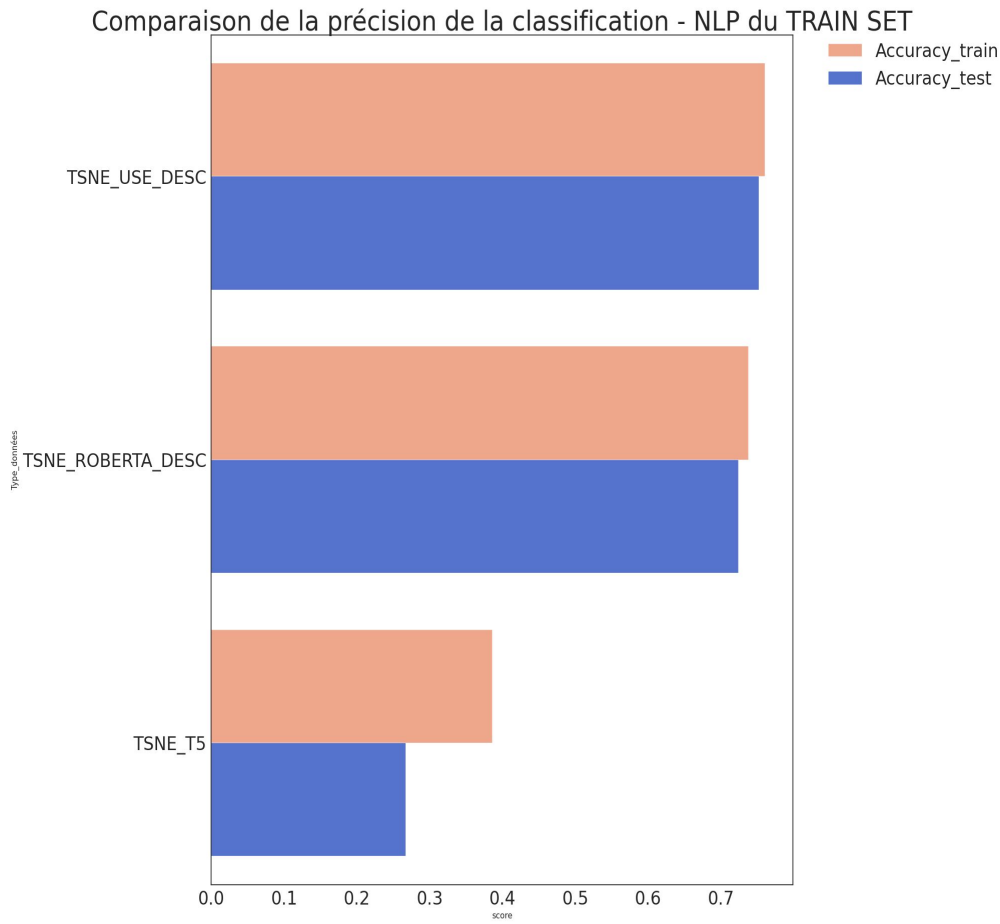
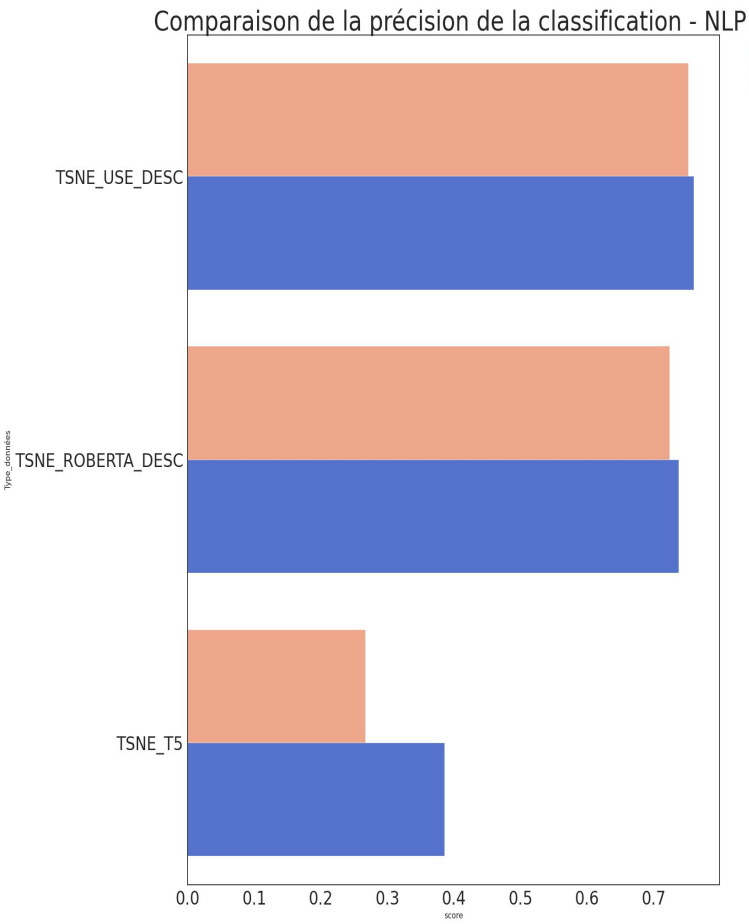
	Model	Accuracy	Recall	Prec.	F1	Kappa	MCC
et	Extra Trees Classifier	0.7607	0.7607	0.7744	0.7604	0.7598	0.7208
rf	Random Forest Classifier	0.7457	0.7457	0.7616	0.7456	0.7032	0.7057
xgboost	Extreme Gradient Boosting	0.7144	0.7144	0.7268	0.7121	0.6667	0.6693
knn	K Neighbors Classifier	0.7117	0.7117	0.7322	0.7097	0.6636	0.6671
lightgbm	Light Gradient Boosting Machine	0.7103	0.7103	0.7198	0.7083	0.6619	0.6641
gbc	Gradient Boosting Classifier	0.6804	0.6804	0.6959	0.6774	0.6271	0.6304
dt	Decision Tree Classifier	0.6790	0.6790	0.6953	0.6791	0.6253	0.6278
qda	Quadratic Discriminant Analysis	0.3239	0.3239	0.3227	0.2923	0.2113	0.2190
nb	Naive Bayes	0.3171	0.3171	0.2936	0.2797	0.2035	0.2119
lr	Logistic Regression	0.2940	0.2940	0.2608	0.2516	0.1762	0.1830
ada	Ada Boost Classifier	0.2870	0.2870	0.2831	0.2591	0.1685	0.1755
ridge	Ridge Classifier	0.2846	0.2846	0.1934	0.2222	0.1652	0.1755
lda	Linear Discriminant Analysis	0.2831	0.2831	0.2377	0.2408	0.1635	0.1698
svm	SVM - Linear Kernel	0.1741	0.1741	0.1583	0.1245	0.0342	0.0414
dummy	Dummy Classifier	0.1361	0.1361	0.0185	0.0326	0.0000	0.0000

	Model	Accuracy	Recall	Prec.	F1	Kappa	MCC
0	Extra Trees Classifier	0.7524	0.7524	0.7494	0.7493	0.7111	0.7116



		TSNE_USE_DESC						
Anciens Modèles	Baby Care	27	3	1	4	7	0	3
	Beauty and Personal Care	3	32	2	4	0	1	3
	Computers	1	2	39	1	0	1	1
	Home Decor & Festive Needs	3	2	4	27	3	5	1
	Home Furnishing	6	0	0	0	38	0	1
	Kitchen & Dining	1	2	2	2	0	35	3
	Watches	0	1	1	2	1	1	39
		0	1	2	3	4	5	6

# Modèle T5 : Comparaisons Résultats NLP



# Conclusion et Suite du Projet

## Dashboard

- ❖ **Dashboard : Mettre en place des ateliers pour le client et que ses remarques soient prises en compte.**
- ❖ **Des améliorations techniques: tailles des graphiques affichées; Charte graphique du client**

## Modele T5

- ❖ **Fournit des résultats satisfaisants malgré l'utilisation de modèle pré-entraîné (T5-small) pour une classification supervisée mais le modele USE avec la version 4 donne de meilleurs résultats**
  - ❖ **Envisager pour la suite, l'utilisation d'un modèle de classification non-supervisée comme K-Means et évaluer leur score ARI.**
- 