Difficulté :

445 compte différente, et des nouveaux comptes à ajouter au fur et au mesure

Si on regroupe les données par fournisseur directement, certains fournisseurs ont très peu de données à apprendre

Les « feature » ne sont pas fortement liée avec le compte

Valeur de TVA, Business\_nature etc sont des valeurs par défaut

Stratégie :

Plan du compte :

Expense :

FRA01 : compte PCI

SHARE : compte E2K (hors Lyxor) AltCompte plus commence par non 9

996- 6xxxxxxx

993- 3xxxxxxxx

SHPCG : compte Lyxor (S4892 et S4893)

Expense Bis :

Compte PCI

996xxxxx :

993xxxxx : compte de charge

PCI commence par 9 BU est filial

Cas particulier :

1. Nouveaux fournisseurs
2. Fournisseur a une seulement compte dans les données historiques
3. Lissage

Features :

|  |  |  |
| --- | --- | --- |
| Feature | Type | Description |
| BUSINESS\_UNIT | Catégoriel | Code entité |
| FILIAL |  | Filale |
| VENDOR\_SETID |  |  |
| VENDOR\_ID |  |  |
| MONETARY\_AMOUNT | Continu |  |
| VAT\_INV\_AMT\_GL |  |  |
| ASST\_FLG |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
| VAT\_TREATMENT |  |  |
| VAT\_SERVICE\_TYPE |  |  |
| VOUCHER\_ID |  |  |
| TAX\_CD\_VAT |  | Code tva |
| DESCR |  | Description de prestation |
| ACTIVE\_FLG |  | Marque de lissage |
| VNDR\_FIELD\_C30\_J |  |  |
| PROB\_VENDOR\_C1 |  |  |
| PROB\_VENDOR\_C3 |  |  |
| PROB\_VENDOR\_C993 |  |  |
| PROB\_VENDOR\_C997 |  |  |
| VENDOR\_ONLY |  |  |
| VENDOR\_ONLY\_993 |  |  |
| VENDOR\_ONLY\_996 |  |  |
|  |  |  |
|  |  |  |
|  |  |  |

Modèle de classification :

1. Modèle de classification du premier chiffre du compte
2. Modèle de classification de 3 premiers chiffres du compte (993, 996, 994 etc)
3. Modèle de classification pour les comptes commencent par 996 :

Traitement par fournisseur

1. Modèle de classification pour les comptes commencent par 993/994/997

Implémentation :

1. Langage de programmation : Python 3.5
2. Algorithme de machine Learning : Random Forest
3. Application : GUI ( Graphical User Interface) et générer un exécutable avec cx\_freeze

Pour distribuer l’outil

Random Forest :

Random Forest est un algorithme de machine learning qui est particulièrement efficace pour repérer des liens entre une variable à expliquer et des variables explicatives. Random Forest va classer les variables explicatives en fonction de leurs liens avec la variable à expliquer.

GUI tkinter :

Tkinter est une boite à outils pour faire des interfaces graphiques en Python. Elle repose sur un autre langage de script, le Tcl (Tool Command Language) en utilisant la bibliothèque Tk de ce langage. l'un des avantages de Tkinter est sa portabilité sur les OS les plus utilisés par le grand public.

Cx\_Freeze :

Cx\_Freeze permet de retrouver avec un fichier .exe et plusieurs fichiers compagnons de fichier .py, bien plus faciles à distribuer et, pour les utilisateurs, à exécuter. De cette façon, on peut transmettre le programme sans obliger les utilisateurs à installer Python sur leur ordinateur.

Avantages de cx\_Freeze

* Portabilité : cx\_Freeze est fait pour fonctionner aussi bien sur Windows que sur Linux ou Mac OS ;
* Compatibilité : cx\_Freeze fonctionne sur des projets Python de la branche 2.X ou 3.X ;
* Simplicité : créer son programme standalone avec cx\_Freeze est simple et rapide ;
* Souplesse : nous pouvons aisément personnaliser le programme standalone avant de le construire.

Evaluation :

1. Matrix de Confusion
2. Temps de calcule

Prédiction du 3 premier chiffre : 2454 lignes 0.444s

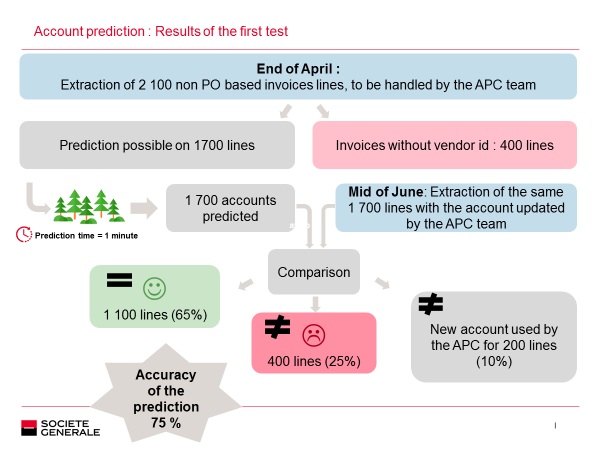
Prédiction du compte 993XXXXX : 0.299 s

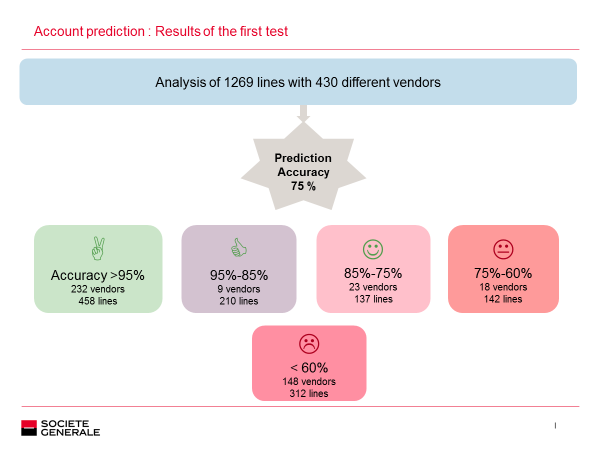
Prédiction du compte 996XXX : 534 fournisseur 120.139 s

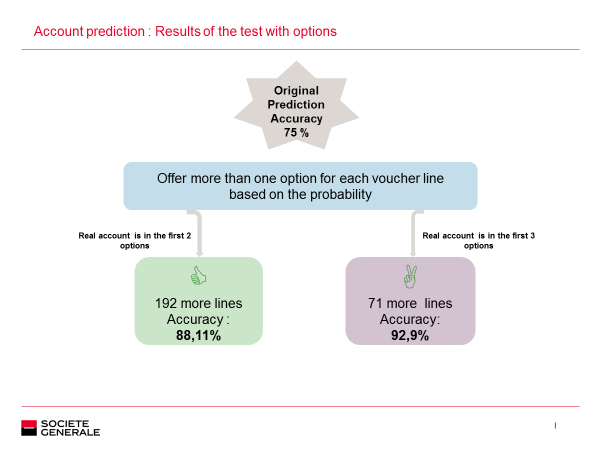
1. Test

Facture\_récyclée d’avril

|  |
| --- |
| Predict the first number : 9 or not  if c1=9:  Filter new vendor  Predict the firt 3 numbers 996,993 or others  filter vendor if only have one account in the historic  if c3=993:  filter vendor if only have one account begin with 993 in the historic  Predict 993XXXX  else if c3=996:  filter vendor if only have one account begin with 996 in the historic  for each Vendor:  predict 996XXXX  Offer the probability of each option(possible account) |







|  |  |  |  |
| --- | --- | --- | --- |
| **Title** | **Précision** | **Erreur** | **Description** |
| **Predicted\_1 :**  **9 or not** | 1834/ 1837    99.84% | 3 lines of Z0010(filial) use the  account 996XXX  Index(968,969,970) | Filtre :  lignes en attant |
| **Predicted\_3 :**  **993 ; 996; autre** | 1437/1480    97.09% | Between 993 et 996 :  35 lines  Prédiction is 993 but real is 996 : 31 lines | Filtre :  NEW\_VENDOR  FILIAL  NEW\_ACCOUNT |
| **Predcited\_all**    **993XXX**  **996XXX(par fournisseur)** | 1112/1480  75.14% |  |  |
| 1109/1437  77.17% |  | Filtre :  Wrong first three numbers |
| 168  100% |  | VENDOR\_ONLY  (168 lines) |
| 941/1269  74.15% |  | Filtrer VENDOR\_ONLY |