



This PDF is generated from authoritative online content, and is provided for convenience only. This PDF cannot be used for legal purposes. For authoritative understanding of what is and is not supported, always use the online content. To copy code samples, always use the online content.

Content Analyzer Plug-in for GAX

[Programmer un entraînement](#)

Programmer un entraînement

Programmez une séance d'entraînement dans l'onglet **Programmation d'entraînement**.

- Pour remplir les champs avec les valeurs d'une séance d'entraînement existante, cliquez sur l'icône de clone. C'est une façon pratique de modifier l'heure à laquelle une séance d'entraînement existante sera exécutée.
- Pour remplir les champs avec des valeurs par défaut, cliquez sur l'icône du signe +.

Dans les deux cas, la boîte de dialogue **Options d'entraînement de modèles** s'affichera.

Model Training Options

* Model Name
Action0015

* Training Object
TrainAct

* Subject Field Treatment
Add to the text

* Training Quality
Regular Level(4)

* Cross Validation
Split to 5 Sets

* Start Time
Oct 28, 2016 2:45:00 AM

* Min Samples in Category
3

* Keyword Threshold
5

* Categories for Training
All Categories

Create

Options d'entraînement de modèles

1. Les **noms de modèles** doivent être composés d'un maximum de 21 caractères et doivent uniquement utiliser les **caractères autorisés**.

2. Traitement du champ Objet

- **Ignorer**—L'entraînement ne prend pas en considération le contenu du champ **Objet**.
- **Ajouter au texte**—L'entraînement prend en considération le contenu du champ **Objet**.
- **Ajouter avec importance double**—L'entraînement donne deux fois plus d'importance au contenu du champ **Objet** qu'à celui du corps du courriel.

3. **Qualité de l'entraînement**—Si vous savez que l'objet de données d'entraînement contient de

nombreux objets texte classés dans les mauvaises catégories, utilisez les Niveaux de non-fiabilité 10-12. Sinon, utilisez Brouillon ou les Niveaux réguliers 1-6. Veuillez noter ce qui suit :

- Les niveaux réguliers et les niveaux de non-fiabilité représentent deux échelles indépendantes qui sont difficilement comparables. Dans chacun de ces types de niveau, un nombre plus grand se traduit par une meilleure qualité. Le seul moyen de confirmer si, par exemple, un niveau de non-fiabilité de 11 produira de meilleurs ou de moins bons résultats qu'un niveau régulier de 4 est de créer un modèle pour chacun des paramètres et de les tester.
 - En fait, ces niveaux permettent de déterminer le nombre de mots que le système prend en considération et le nombre d'itérations que le processus de traitement exécute. Augmenter chacun de ces niveaux devrait améliorer la qualité du modèle résultant, mais ce ne sera pas le cas à des niveaux plus élevés. Encore une fois, la seule façon de s'en assurer est de tester les modèles résultants, préféablement par **validation croisée**.
 - La durée de l'entraînement augmente à mesure que vous passez de la qualité Brouillon à celle d'un Niveau régulier de 3. Toutefois, après avoir dépassé une qualité de niveau 3, la différence en matière de durée de l'entraînement est minime. Genesys vous recommande de seulement utiliser le plus bas niveau de qualité lorsque vous souhaitez obtenir une lecture préliminaire de l'estimation de la qualité du modèle. Pour la production, utilisez un niveau de qualité de 2-6.
4. La **validation croisée** est expliquée sur une autre page. Choisissez entre aucune validation croisée ou une validation croisée qui séparera les données en 3, en 6 ou en 10 jeux. Si vous sélectionnez la validation croisée, l'entraînement produira une mesure de précision pour le modèle ainsi que le modèle lui-même. Cela a l'avantage de ne nécessiter aucune autre étape d'essai, mais la durée de l'entraînement sera plus longue.
 5. Étant donné que l'entraînement peut utiliser une grande partie des ressources du système, il est préférable de programmer **l'heure de début** à une heure non achalandée. Assurez-vous de programmer une heure de début à un moment postérieur au moment présent.
 6. Le **nombre minimal d'échantillons dans la catégorie** correspond au nombre minimal d'objets texte qu'une catégorie doit avoir pour qu'elle soit comprise dans l'entraînement. Les catégories qui ne contiennent pas ou peu d'objets texte ne sont pas pertinentes pour l'entraînement.
 7. Le **seuil de mots-clés** correspond au nombre minimal d'objets texte dans lesquels un mot-clé doit apparaître afin que celui-ci soit pris en considération dans l'entraînement. Une valeur relativement élevée de ce paramètre peut réduire la durée de l'entraînement, mais aussi la qualité. Ce qui est considéré comme étant de valeur élevée ou faible pour ce paramètre dépend de la taille totale de l'objet de données d'entraînement. Par exemple, si un objet de données d'entraînement contient de 5 à 10 objets texte par catégorie, un seuil de mots-clés élevé correspondrait à **2** ou à **3**. En revanche, si un objet de données d'entraînement contient de 30 à 50 objets texte par catégorie, un seuil de mots-clés élevé correspondrait à **20**.
 8. Les **catégories pour l'entraînement** sont **Toutes les catégories** ou **Catégories terminales seulement**. Le terme "catégorie terminale" se dit d'une catégorie qui ne contient pas de sous-catégorie. Il est possible qu'une arborescence de catégories utilise des catégories non terminales principalement pour organiser celles qui sont terminales. Si c'est le cas, peu ou pas d'objets texte sont associés aux catégories non terminales, donc il y a peu à gagner à ajouter ces dernières à l'entraînement.
 9. Vous pouvez nettoyer votre objet de données d'entraînement en utilisant la **sous-fenêtre de prétraitement du texte** (à la droite de **Options d'entraînement de modèles**) pour éliminer le texte superflu.