



# CHAIRE D/CC

## Données/Connaissance Client

# TOPCAPY

TRAITEMENT OPÉRATIONNEL DE LA PAROLE CLIENT AVEC PYTHON

## Guide de prise en main de l'outil

### TOPCAPY en quelques mots

**TOPCAPY** est un outil d'aide à l'analyse de la parole client, développé par la chaire Données / Connaissance Client, hébergée par la fondation Université Bretagne Sud.

**TOPCAPY** est une application visuelle et interactive, développée en Python. Elle permet de réaliser aisément des modélisations thématiques automatiques d'un corpus de commentaires (via des modèles LDA - *Latent Dirichlet Allocation*), accompagnée d'une analyse exploratoire visuelle, reposant notamment sur un élément de notation (score NPS, ou autre).

**TOPCAPY** est composé de deux modules :

- **TOPCAPY-model**, pour réaliser les modélisations LDA ;
- **TOPCAPY-exploration**, pour explorer les résultats ainsi obtenus.

### Prérequis :

L'utilisation de **TOPCAPY** et l'exploitation des résultats associés nécessitent des connaissances dans le domaine du machine learning, et plus particulièrement du Traitement Automatique du Langage Naturel (TALN). En particulier, **la bonne compréhension du processus de modélisation LDA est un prérequis indispensable.**

Les données analysées doivent être constituées d'un **champ de commentaires clients**, ainsi que d'un **champ de notation**.

La convergence des algorithmes nécessite une puissance et un temps machine importants, à adapter en fonction de la structure des commentaires (notamment en fonction de leur hétérogénéité) et de leur volume.

### Phase 1 : répertoire de travail et paramétrages

La procédure de mise en place des répertoires de travail et du paramétrage est la suivante :

- copier le répertoire **TOPCAPY** à l'endroit souhaité
- placer les données à analyser dans le répertoire **data** (format csv, séparateur ';')
- dans le répertoire **TOPCAPY** se trouvent deux fichiers de paramétrage : **topcapy\_parameters\_grade.csv** et **topcapy\_parameters\_app.csv**.

- dans **topcopy\_parameters\_grade.csv** sont paramétrés les liens entre les notes et les statuts, ainsi que les couleurs des graphiques associés. Les paramètres pré-enregistrés sont relatifs à une notation NPS.
- dans **topcopy\_parameters\_app.csv** doivent être renseignés le nom du fichier à analyser (**file\_name**), le nom de la variable contenant les verbatims (**var\_name**) et le nom de la variable de notation (**grade\_name**).

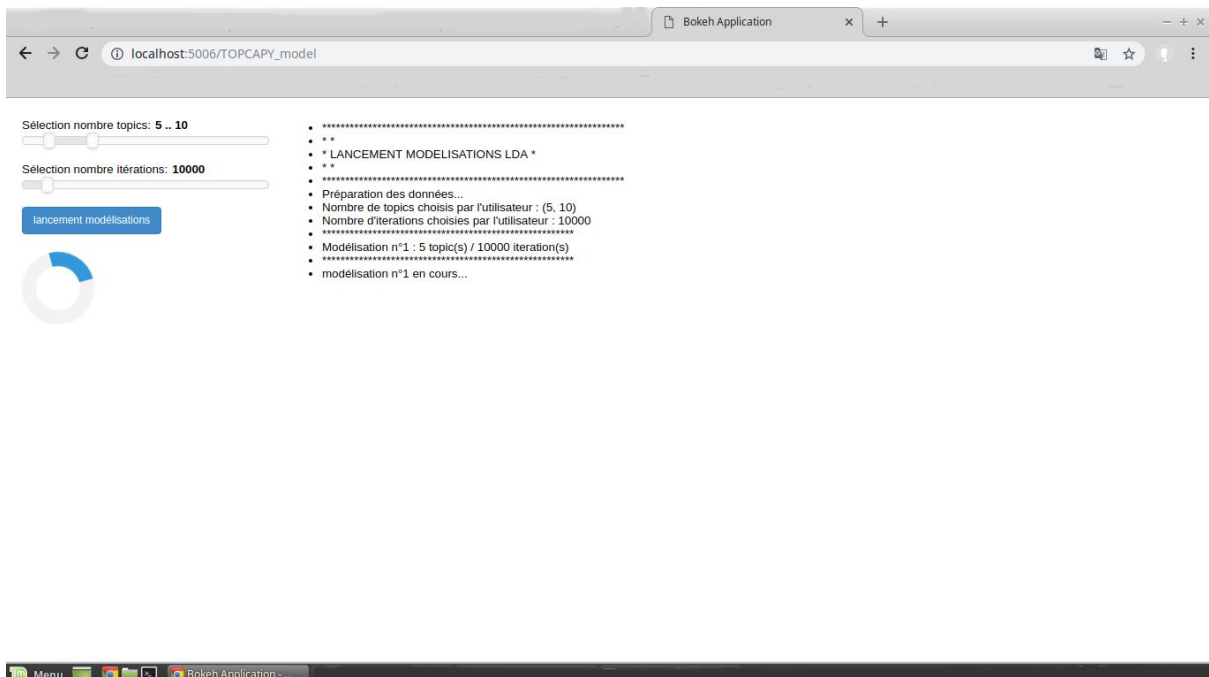
## Phase 2 : modélisation

Le module **TOPCAPY-model** permet de tester plusieurs modélisations thématiques de type LDA sur le corpus textuel. La procédure à suivre est la suivante :

- se placer dans le répertoire TOPCAPY en console
- lancer un serveur bokeh : **bokeh serve TOPCAPY\_model**
- l'applicatif relatif à la modélisation est alors accessible *via* l'url :  
**http://localhost:5006/TOPCAPY\_model**



- choisir le nombre de thèmes *a priori*, et le nombre d'itérations de l'algorithme. L'évolution des traitements est affichée dans l'application :



- une fois les modélisations terminées, les caractéristiques des modèles sont automatiquement enregistrés dans le répertoire "model" et peuvent être analysés *via* les visualisations interactives (sorties sous format html dans ce même répertoire).

### Phase 3 : analyse exploratoire

L'objectif de l'analyse exploratoire proposée par le module **TOPCAPY-exploration** est d'analyser la structure thématique du corpus de commentaires clients et d'identifier ainsi les sources de satisfaction et d'insatisfaction à travers le vocabulaire utilisé, la ponctuation, les structures des phrases.

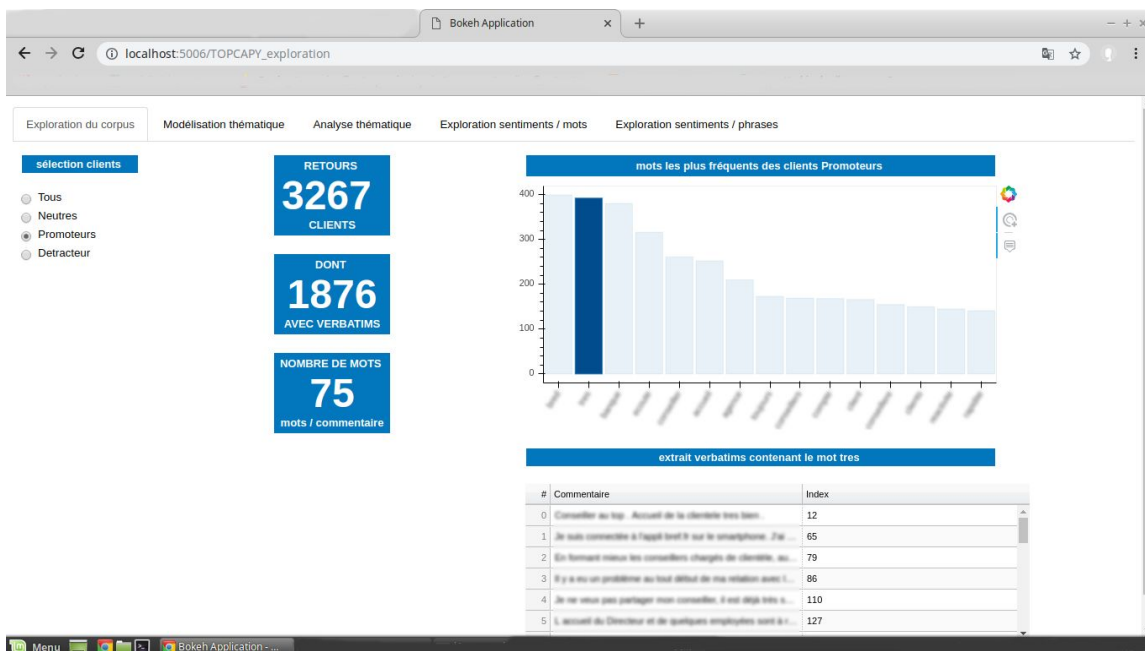
Pour cela il est nécessaire de disposer d'un jeu de données contenant un élément de notation client (note pour score NPS, ou autre échelle de notation).

La procédure est la suivante :

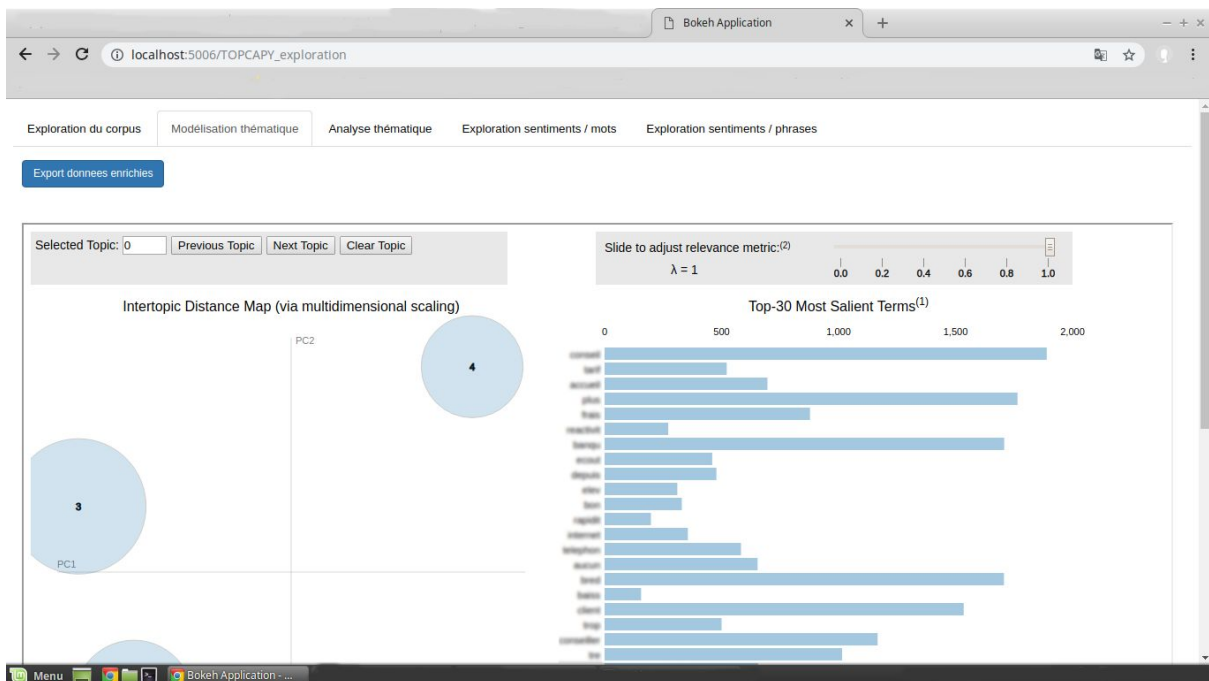
- indiquer le nom de la modélisation retenue dans le fichier de paramétrage **topcopy\_parameters\_app.csv**.
- lancer le serveur bokeh associé : **bokeh serve TOPCAPY\_exploration** et accéder à l'appliquatif *via* l'url **http://localhost:5006/TOPCAPY\_exploration**

NB : lors du premier lancement de **TOPCAPY-exploration** avec une nouvelle modélisation et de nouvelles données, un fichier de "vocabulaire" sera automatiquement créé. Cette création peut être longue en fonction de la richesse des données d'entrée : son évolution sera suivie *via* le terminal. Par la suite, lorsque **TOPCAPY-exploration** sera utilisé avec ces mêmes données, cette étape ne sera pas rejouée (fichier de vocabulaire automatiquement sauvegardé).

- l'onglet "**exploration du corpus**" fournit une analyse exploratoire (nombre de lignes analysées, nombre de verbatims, nombre de mots par commentaire) en fonction du statut de la notation sélectionné, ainsi qu'une extraction de commentaires clients.

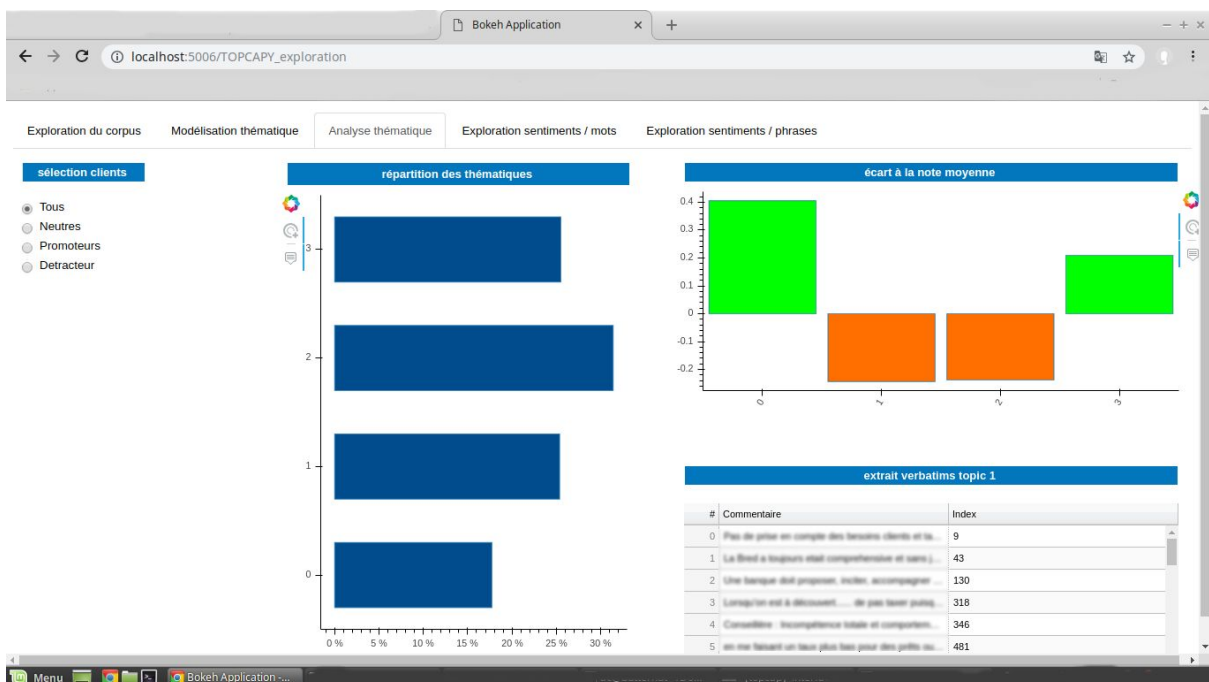


- l'onglet “**modélisation thématique**” fournit une visualisation interactive de la modélisation retenue (mots représentatifs pour chaque classe) :

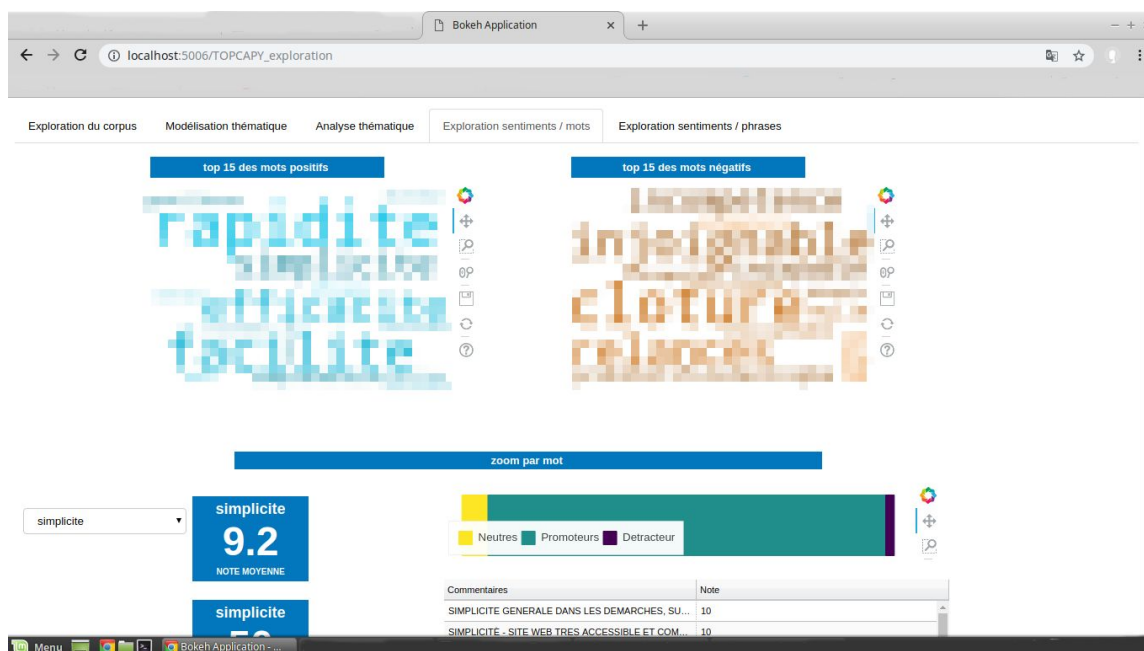


En cliquant sur le bouton “Export données enrichies”, l'utilisateur a la possibilité d'exporter les données enrichies des résultats de modélisation. Le fichier **data\_with\_model\_MODELNAME.csv** ainsi obtenu est enregistré dans le répertoire **data**.

- l'onglet “**analyse thématique**” permet une visualisation de la répartition des thématiques dans le corpus, ainsi qu'une analyse conjointe thématique / note, accompagnée d'une extraction de verbatims.



- l'onglet "**exploration sentiments / mots**" permet d'effectuer une analyse de sentiments du corpus basée sur le vocabulaire utilisé. Les mots les plus représentatifs des sentiments négatifs et positifs sont représentés sous la forme de *wordclouds*. L'utilisateur peut effectuer un zoom par mot mettant en avant la note moyenne des commentaires contenant le mot sélectionné, et une extraction des commentaires clients concernés.



- l'onglet "**exploration sentiments / phrases**" permet d'effectuer une analyse des sentiments clients sur la base de la structure des phrases employées : longueur du verbatim, ponctuation.

