

# Projet VBA :

## *Cahier des charges*

### *Labélisation automatique des données texte*

Réalisé par :

- Soukaina CHAKOUR
- Abdelmoula ELBOUHALI
- Taha LAKHDARI

Encadré par :

- Mme. Selwa EL FIRDOUSSI

# *Introduction*

Ce projet s'inscrit dans le cadre du cours d'informatique - Visual Basic for Application (VBA) destiné à la deuxième année cycle ingénieur de l'Emines. Ceci est l'occasion pour nous de mettre en pratique les connaissances acquises durant ce cours, et de manifester notre esprit de créativité.

Dans ce projet, nous sommes appelés à mener à bien pour une durée de deux mois un projet sous VBA qui a pour finalité la labélisation automatique des données de format texte.

## *1. Contexte générale :*

Le projet servira comme solution qui répond aux besoins de l'université en premier lieux et aux besoins des entreprises et individus en second lieu. La solution de Développement d'un outil sous VBA permettant la labélisation automatique des données sous format de texte, permettra à partir d'une base de données, sous format texte, d'effectuer des opérations de catégorisation ou de classification selon un critère précis.

## *2. Problématique :*

L'annotation et la catégorisation des données sont des tâches difficiles à effectuer manuellement en posant des caractéristiques et des critères précises avec des grandes base de données. En fait, cette labélisation manuelle pose un vrai problème de temps, de marge d'erreur et de traitement des textes inutiles. A titre d'exemple, dans les grandes entreprises, le choix des bons CV parmi ceux qui ont postulé à une offre d'emploi peut prendre une semaine ou plus, aussi une détection automatique des commentaires offensives sur une plateforme pour les supprimer ou bloquer leurs éditeurs, est un besoin réel auquel il est impossible d'opter pour une solution manuelle.

## *3. Définition du besoin :*

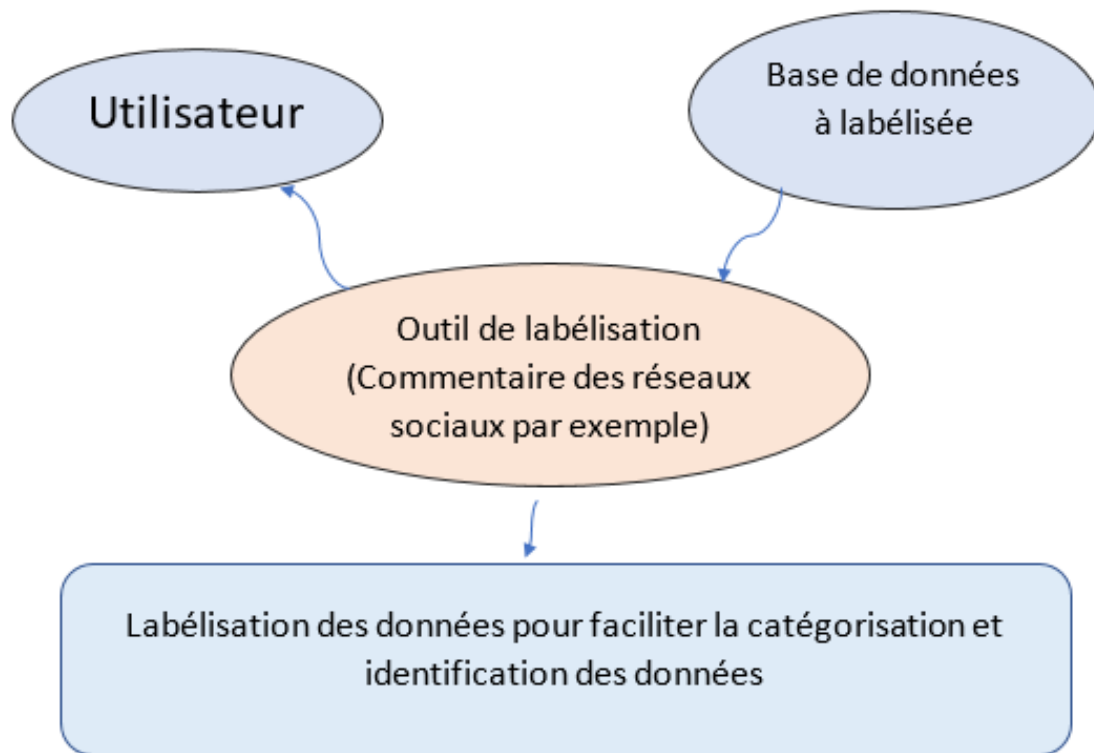


Figure 1:Diagramme bête à cornes

#### 4. Contraintes :

- *Collecte des données : La difficulté de collecter les données par la méthode de scrapping.*
- *Recherche des meilleurs algorithmes pour chaque fonctionnalité sous VBA.*
- *Généralisation de la première étape du projet sur tous les types de la data.*
- *Coding : la multiplicité et le nombre des fonctionnalités demande un algorithme et un code pour chacune.*
- *Difficulté de détection des erreurs avec le debugger VBA.*

## *5. Solution proposée :*

Pour répondre à la problématique détaillée auparavant, nous sommes invités à concevoir un outil VBA qui permettra de labéliser des données de format texte sous un nombre déterminé de catégories. En fait, l'utilisateur sera amené à fournir son fichier de données, à définir sa liste de catégories de labélisation souhaitée, et à préciser les mots clés définissant chaque catégorie.

Cependant, comme première étape de notre réalisation de l'outil on se contentera de développer une version statique où nous travaillerons sur une base de données des commentaires de réseaux sociaux et nous essayerons de la labéliser sous deux catégories : offensive et inoffensive. Et pour collecter cette base de données, nous allons recourir au scrapping des commentaires d'un réseau sociaux sur un sujet, une personnalité ou une marque commerciale.

## *6. Fonctionnalités de l'outil :*

- Saisie des données :

Au démarrage de l'application, l'utilisateur aura un User Forme où il télécharge son fichier de données sous format Excel et saisie les catégories de labélisation avec les mots clés de chacune de celles-ci dans les champs de texte dédiés.

- Génération de la labélisation :

En soumettant le User Forme, le fichier Excel fournis est régénéré à l'utilisateur en ajoutant à chaque ligne la labélisation convenable.

- Visualisation des données :

Il s'agit d'un Dashboard présentant des statistiques des groupes de labélisation présents dans la base de données.