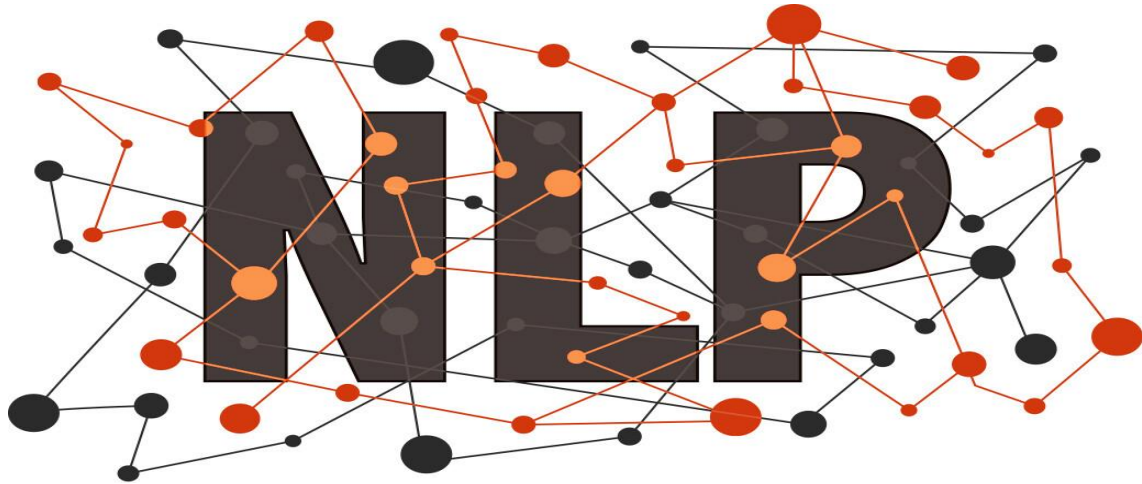




كلية العلوم الرباط  
Faculté des sciences Rabat



جامعة محمد الخامس بالرباط  
Université Mohammed V de Rabat



# RAPPORT DU PROJET

## Analyse des sentiments

Réalisé par :

**FADOUA BOUHOU**  
**SOUMIA BENYAKKOU**

Encadré par :

**Mr. ABDELHAK MAHMOUDI**

**Master Spécialisé Ingénierie de Données et Développement Logiciel**

**ANNEE ACADEMIQUE : 2020/2022**



## Sommaire :

1. INTRODUCTION.....	3
2. NATURAL LANGUAGE PROCESSING .....	4
2.1. CONCEPT .....	4
2.2. UTILISATION .....	4
3. ANALYSE DES SENTIMENTS .....	5
3.1. DEFINITION .....	5
3.2. EXEMPLE .....	5
4. IMPLEMENTATION .....	5
4.1. OUTILS .....	5
4.2. TEXTBLOB .....	6
4.3. CODE.....	6
5. TEST .....	8
5.1. SCENARIO 1 – TEXTE .....	8
5.2. SCENARIO 2 - VIA UN FICHIER .....	9
5.3. SCENARIO 3 - VOIX .....	10
6. CONCLUSION .....	11

## 1. INTRODUCTION



Le présent rapport résume notre période de réalisation du ce programme, le programme est d'implémenter d'une application pour l'analyse des sentiments (via un texte donné, document ou une voix)

En effet la réalisation d'un projet informatique est avant tout un projet qui doit être écrit, planifié et surtout évalué. Il s'agit de se fixer des objectifs, et aussi définir des étapes. Lorsque l'on souhaite réaliser une application, la première des choses est de définir le projet. L'aperception d'un projet est essentielle pour sa réussite.

Ce rapport décrit le projet et son contexte et les phases nécessaires à sa réalisation. Ainsi, que les data sets en générale.

Grâce à cette expérience pratique, nous avons eu l'opportunité de découvrir :

- ✓ Natural Language Processing « NLP »
- ✓ Analyse des sentiments
- ✓ La reconnaissance vocale
- ✓ La bibliothèque TextBlob

## 2. NATURAL LANGUAGE PROCESSING

### 2.1. CONCEPT

**Le traitement Automatique de la parole** a principalement deux domaines :

- ✓ La reconnaissance automatique de la parole.
- ✓ La synthèse vocale.

**La reconnaissance automatique de la parole** est une branche très importante du Machine Learning et donc de l'intelligence artificielle. Permet à la machine de comprendre et de traiter des informations fournies (oralement) par un utilisateur humain.

Elle consiste à employer des techniques afin de comparer une onde sonore à un ensemble d'échantillons, composés généralement de mots.

**La synthèse vocale** est l'ensemble des dispositifs, matériels ou algorithmes, pour générer automatiquement de la parole artificielle. La synthèse de la parole consiste en la lecture par une voix synthétique d'un texte numérique.

Il s'agit simplement de prononcer un petit nombre de courtes phrases préenregistrées pour des dispositifs de réponse vocale.

### 2.2. UTILISATION

Des applications basées sur **Le traitement Automatique de la parole** :



Ils existent plusieurs applications dans le domaine l'assistance tels que : Siri et Cortana



Ils existent plusieurs applications dans le domaine de Traduction tels que : **DeepL** et **Google Traduction**



Il est utile aussi pour la détection des spams

## 3. ANALYSE DES SENTIMENTS

### 3.1. DEFINITION

Analyse des sentiments (aussi appelé opinion mining) : consiste à déterminer si l'attitude d'une personne à propos d'un sujet ou du produit est positive, négative ou neutre, aussi de savoir le ressenti général d'une communauté.

Ce procédé apparaît au début des années 2000 et connaît un succès grandissant dû à l'abondance de données provenant de réseaux sociaux, notamment celles fournies par **Twitter**.

### 3.2. EXEMPLE



Neutre

Python ou le R n'ont pas d'importance pour moi, je maîtrise les deux.



Python est magnifique ! j'aime la simplicité.



Positive



Négative

J'aime Python, mais je déteste le R.





## 4. IMPLEMENTATION

### 4.1. OUTILS

Pour développer l'application, nous avons utilisé :



**Python** : est un langage de programmation interprété, multi-paradigme et multiplateformes. Il favorise la programmation impérative structurée, fonctionnelle et orientée objet. Il est doté d'un typage dynamique fort, d'une gestion automatique de la mémoire par ramasse-miettes et d'un système de gestion d'exceptions.

	<p><b>Anaconda</b> : est une distribution des langages de programmation Python et R appliqué au développement d'applications dédiées à la science des données et à l'apprentissage automatique (traitement de données, analyse prédictive, calcul scientifique), qui vise à simplifier la gestion des paquets et de déploiement.</p>
	<p><b>Spyder</b> : (nommé Pydee dans ses premières versions) est un environnement de développement pour Python multiplateforme (Windows, Mac OS, GNU/Linux), il intègre de nombreuses bibliothèques d'usage scientifique : <b>Matplotlib, NumPy, SciPy et IPython</b>.</p>

#### 4.2. TEXTBLOB

**TextBlob** : est une bibliothèque Python pour le traitement de données textuelles. Elle fournit une API simple pour accéder à ses méthodes et effectuer des tâches de traitement du langage naturel (NLP) telles que le balisage de parties de discours, l'extraction de phrases nominatives, l'analyse des sentiments, la classification, la traduction, ...

Pour l'utiliser, il faut d'abord sur Windows ouvrir command prompt et exécuter la commande ci-dessous :

```
pip install -U textblob
```

L'importer et crée un objet textblob pour le manipuler :

```
from textblob import TextBlob
Blob = TextBlob('I love python')
Blob.sentences.words          # [['I', ',', "'love'", ',', "'python'"]]
Blob.tags                     # [ ('I', 'PRP'), ('love', 'VBP'), ('python', 'NN')]
Blob.sentiment.polarity       #Polarity:0.5
Blob.sentiment.Subjectivity   #Subjectivity:0.6
Blob.detect_language()        #'en'
Blob.translate(to = 'fr')     #'j'aime python'
TextBlob('helps').singularize() #help
w = Word('running')
w.lemmatize("v")              #'run'
```

#### 4.3. CODE

```
1 # Core Packages
2 from textblob import TextBlob
3 import tkinter as tk
4 from tkinter import *
5 from tkinter import ttk
6 from tkinter.scrolledtext import *
7
8
9
10 # NLP Packages
11 import nltk
12 import spacy
13 nlp = spacy.load('en')
14 import speech_recognition as sr
15 import pyttsx3
16
17 nltk.download('punkt')
18 nltk.download('averaged_perceptron_tagger')
19
```

L'application nécessite l'installation des bibliothèques suivantes :

**b** : la bibliothèque graphique libre d'origine pour le langage Python, permettant la création d'interfaces graphiques.

**SpeechRecognition** : la bibliothèque pour effectuer la reconnaissance vocale

**TextBlob** : la bibliothèque Python pour le traitement de données textuelles.

**SpaCy** : la bibliothèque open-source gratuite pour le traitement du langage naturel en Python

**NLTK** : une suite de bibliothèques et de programmes pour le traitement symbolique et statistique du langage naturel.

Comment installer les packages ?

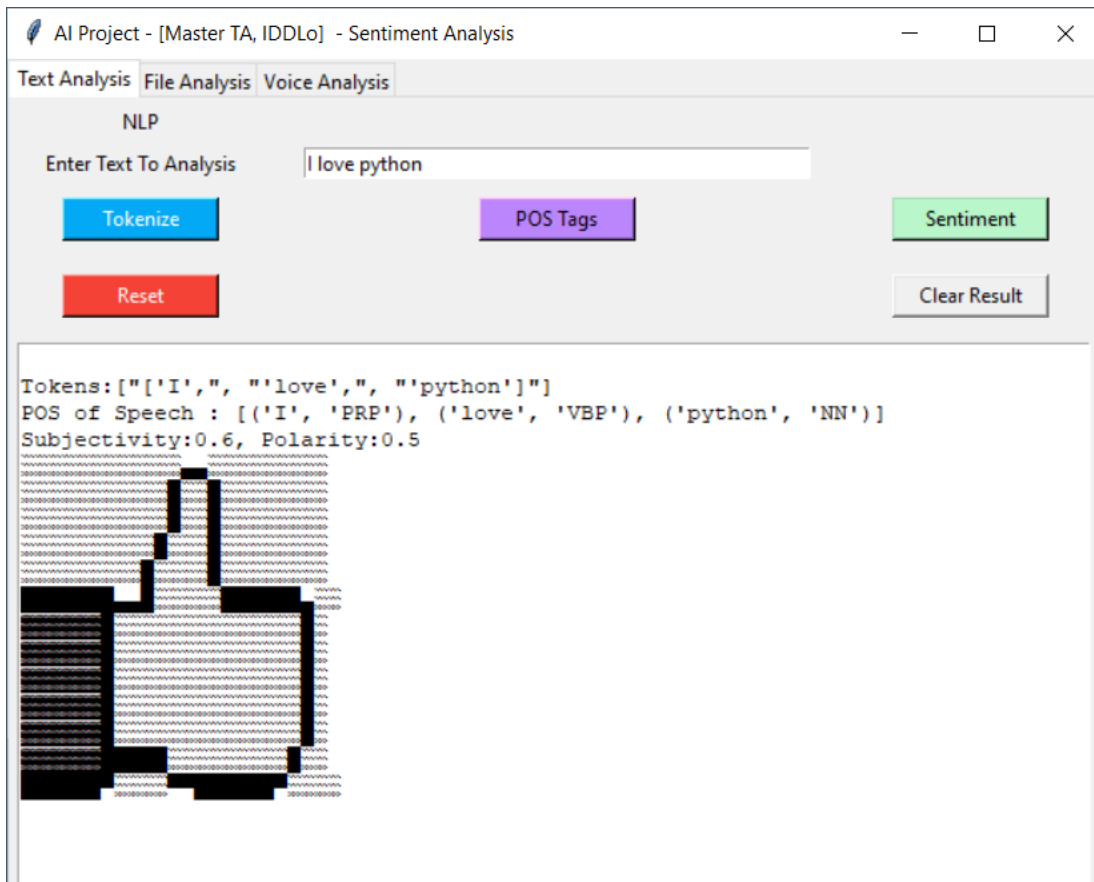
Sur Windows ouvrir la command prompt et exécuter les commandes ci-dessous :

Merci de noter qu'il faut installer au début la version **64 bit** du Python

```
#####  
#Installing Dependencies#  
#####  
# install Python 64 bit version  
  
python -m pip install --upgrade pip  
Pip install numpy  
Pip install nltk  
pip install textblob  
pip install spacy  
python -m spacy download en  
  
pip install pipwin  
pipwin install pyaudio  
pip install SpeechRecognition  
pip install pyttsx3  
pip install tkinter
```

## 5. TEST

### 5.1. SCENARIO 1 – TEXTE



AI Project - [Master TA, IDDL0] - Sentiment Analysis

Text Analysis | File Analysis | Voice Analysis

NLP

Enter Text To Analysis: I love python

Buttons: Tokenize, POS Tags, Sentiment, Reset, Clear Result

Output:

```
Tokens:[['I', ',', 'love', ',', 'python']]  
POS of Speech : [('I', 'PRP'), ('love', 'VBP'), ('python', 'NN')]  
Subjectivity:0.6, Polarity:0.5
```



Le premier onglet offre la fonctionnalité d'analyser les sentiments et d'appliquer le NLP sur un texte saisi par l'utilisateur.

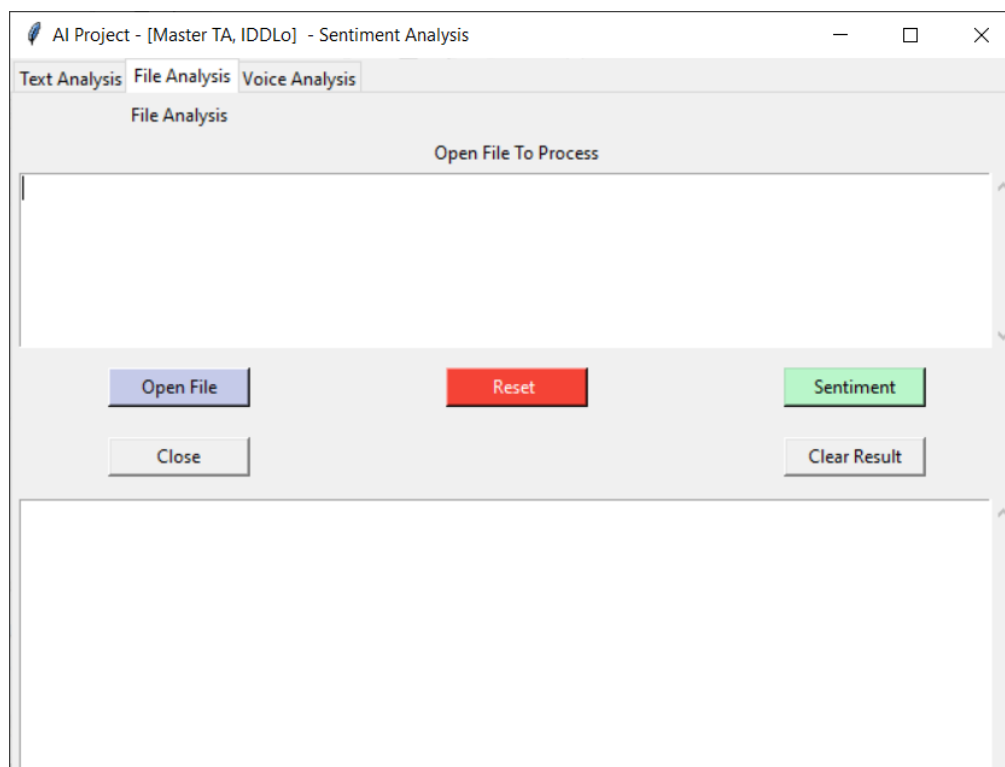
Le bouton Tokenize fait référence à la division d'un texte ou d'une phrase en une séquence de jetons, qui correspondent à peu près à des « mots ». C'est l'une des tâches de base de la NLP.

Le bouton POS Tags : le balisage grammatical est une méthode pour marquer les mots présents dans un texte sur la base de sa définition et de son contexte

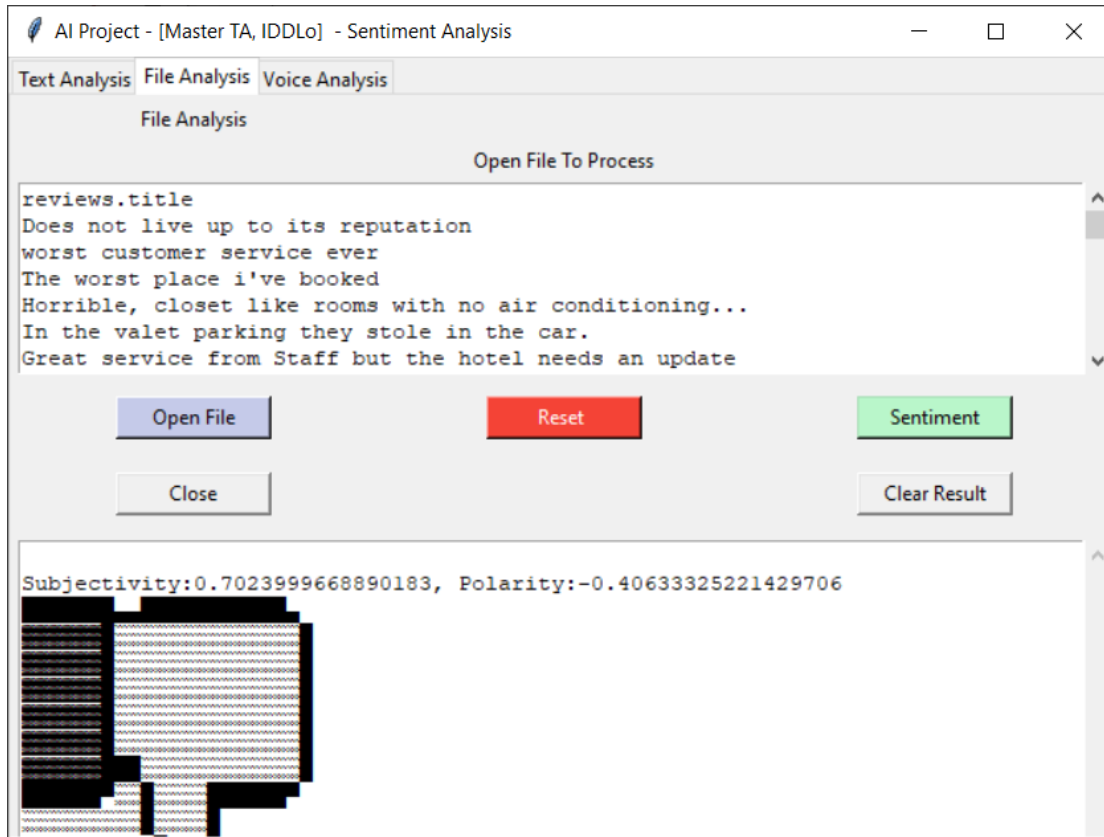
Le bouton Sentiment : est essentiellement le processus de détermination de l'attitude ou de l'émotion de l'écrivain, c'est-à-dire s'il est positif ou négatif ou neutre.

La polarité est un flottant qui se situe dans la plage de  $[-1,1]$  où 1 signifie une déclaration positive et -1 signifie une déclaration négative. Les phrases subjectives se réfèrent généralement à une opinion personnelle, à une émotion ou à un jugement tandis que l'objectif fait référence à des informations factuelles. La subjectivité est également un flotteur qui se situe dans la plage de  $[0,1]$

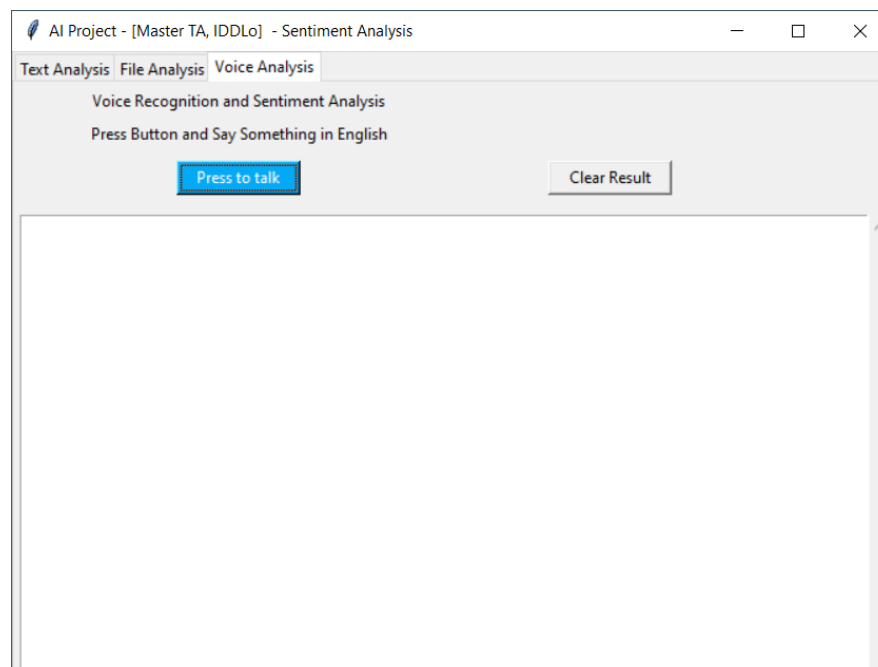
## 5.2. SCENARIO 2 - VIA UN FICHIER



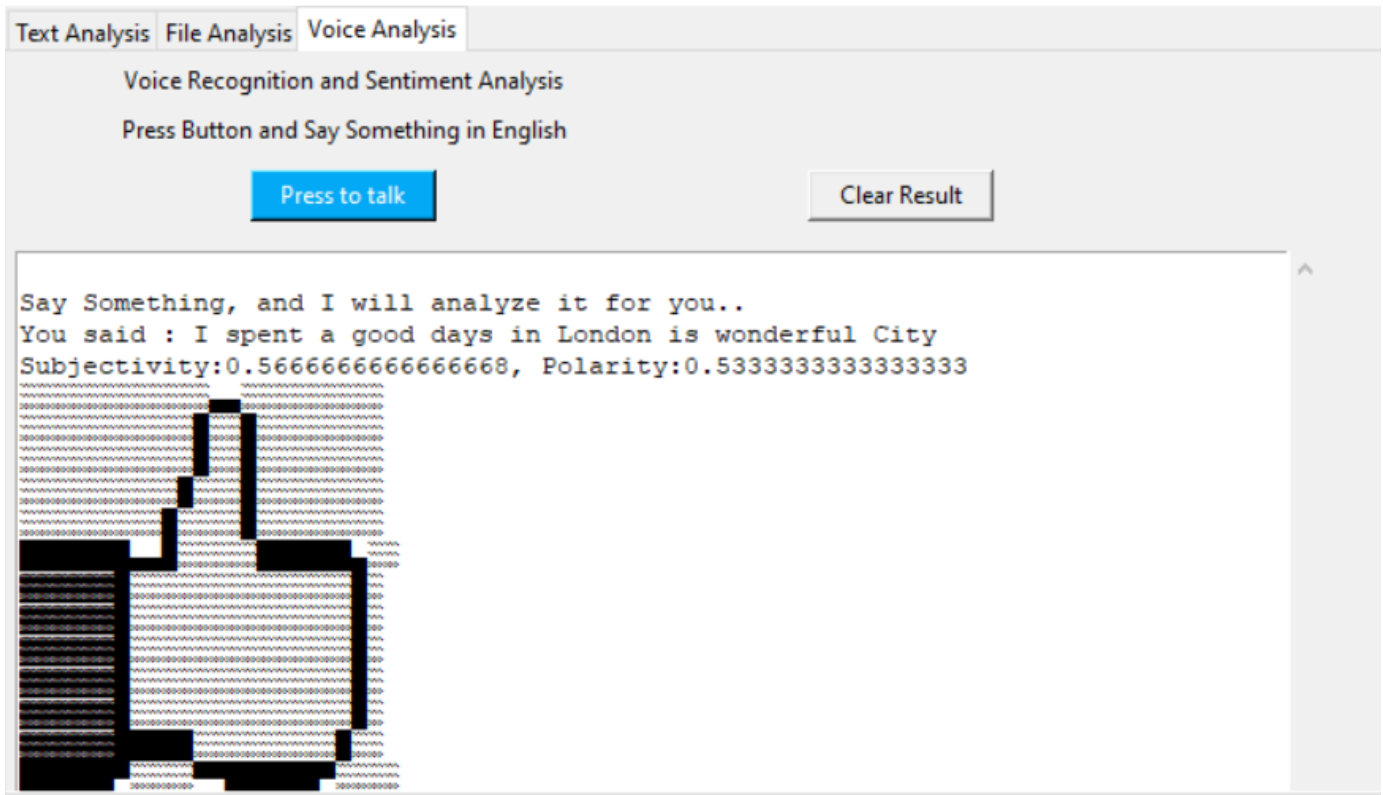
L'onglet File Analysis permet d'analyser les sentiment et l'émotion depuis un fichier texte, le fichier peut contenir un ensemble des feedbacks des clients sur un produit ou service.



### 5.3. SCENARIO 3 - VOIX



L'onglet Voice Analysis permet d'analyser l'émotion et l'attitude depuis un discours vocal en temps réel, en cliquant sur le bouton « **Press to talk** » le micro s'active et le programme affiche ce que l'utilisateur a dit, ainsi que le résultat d'analyse des sentiments.



L'onglet Voice Analysis permet d'analyser l'émotion et l'attitude depuis un discours vocal en temps réel, en cliquant sur le bouton « Press to talk » le micro s'active et le programme affiche ce que l'utilisateur a dit, ainsi que le résultat d'analyse des sentiments.

## 6. CONCLUSION

IBM a bâti son offre de NLP à partir d'**AlchemyAPI**, technologie qu'il a acquise en 2015 et intégrée la même année à son offre cloud Watson. Elle a été rebaptisée pour l'occasion **Watson Natural Language Understanding (Watson NLU)**. De la catégorisation de contenus à l'analyse de sentiments en passant par la détection de mots-clés, ce qui a donné beaucoup de visibilité à Watson NLU.