Projet Informatique Individuel

de Beasse Joseph

DeepLanguage

Cahier des charges

I- Introduction

- Objectif du projet
- Contexte et justification
- Énoncé du problème

II- Méthodologie

- Choix et description des données
- Réalisation techniques: Machine Learning
- Cas d'utilisation
- Outils utilisés

III- Planification

IV- Conclusion

- Résultats attendus
- Limitations et améliorations possibles

I - Introduction

Objectif du projet

L'objectif de ce projet est de développer un modèle de machine learning capable de reconnaître la langue parlée dans un enregistrement ou un fichier audio.
Ce dernier se place dans la catégorie de l'apprentissage supervisé ou plus particulièrement de la classification, ici d'audios.
À terme, le modèle sera hébergé en ligne et fonctionnera sur au moins 4 langues.

Contexte et justification

L'émergence des IA lors de cette dernière décennie a été fulgurante. Ce projet me permettra d'acquérir des compétences dans un domaine particulièrement en vogue: l'intelligence artificielle.
Les modèles permettant de classifier et de reconnaître des "pattern" dans les images sont omniprésents dans notre vie quotidienne.
Acquérir des compétences dans ce domaine et les coupler avec des techniques de traitement de signal font de ce projet une initiation complète aux enjeux que peut rencontrer une entreprise.

Enoncé du problème

Comment créer un modèle de reconnaissance qui peut identifier efficacement la langue dans des enregistrements audio?

II - Méthodologie

Choix et description des données

La récupération des données utiles à l'entraînement du modèle se fera exclusivement à l'aide de la base de données libre de droit nommée "Common Voice", fournie par Mozilla.
Les données sont des enregistrements de quelques secondes de personnes volontaires. Les enregistrements sont soumis par des volontaires du monde entier qui lisent des phrases sélectionnées à haute voix.
Pour réduire la complexité et le temps d'entraînement du modèle, seulement 4 langues seront sélectionnées: Français, Anglais, Allemand et Espagnol.
Le nombre d'enregistrements nécessaires est à déterminer ultérieurement.
Pour maximiser les résultats il faut en effet disposer du plus grand échantillon possible. Devoir ré-entrainer un modèle est plus coûteux en temps que de sélectionner plus de données au préalable.

Réalisation techniques: Machine Learning

Les avantages du machine learning sont nombreux comme entre autres la reconnaissance des patterns “invisibles à l'œil nu’’ dans des données.
Pour entraîner un modèle, on ne peut présenter que des données chiffrées, sous forme de tenseur. Pour cela, on peut extraire différentes informations de chaque fichier audio. Les plus importantes sont: l'allure du spectre audio et les MFCC *(Mel-frequency cepstral coefficients)*.
Une fois ces features sélectionnées, il faudra trouver un modèle cohérent. Une première approche serait d’utiliser des réseaux de neurones convolutionnels (CNN). Ces derniers se sont montrés prometteurs dans de nombreuses tâches de reconnaissance de la parole.
Ensuite vient l'entraînement du modèle en utilisant un algorithme d'optimisation (de type: SGD) puis l’évaluation du modèle et enfin de potentiels ajustements ou son déploiement.

Cas d’utilisation

L’outil peut être utilisé à des fins personnelles.

1. Reconnaître la langue que parle un passant dans la rue.
2. Identifier la provenance d’un interlocuteur dans un enregistrement audio.

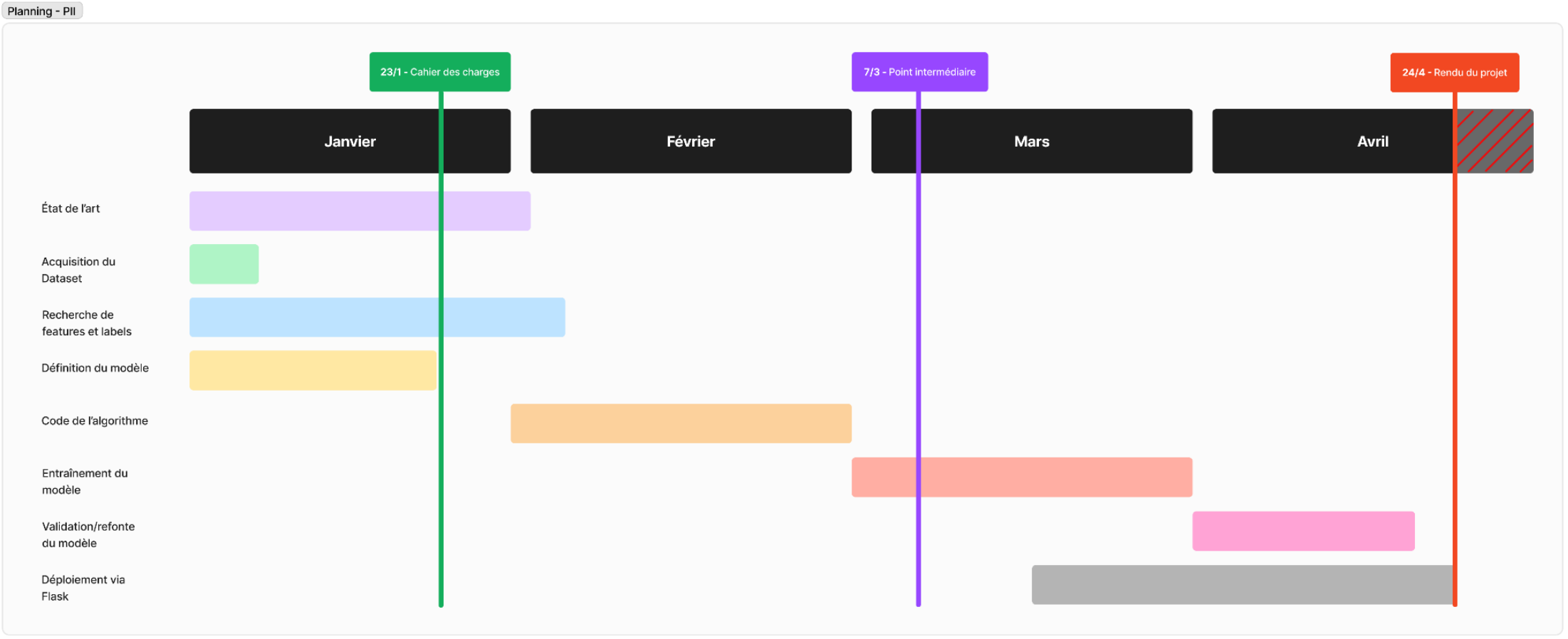
Mais également professionnelles.

1. Application de transcription automatique.
2. Application de traduction en temps réel.
3. Application de reconnaissance vocale.
4. Application de sous-titrage automatique.

Outils utilisés

Pour réaliser ce projet, le langage **Python** sera sollicité pour la réalisation du modèle.
Le framework **Flask** permet lui d’implémenter ce code pour l’utiliser sur un site web.
Le code sera versionné sur la plateforme **Github** où des commits réguliers seront effectués.
Pour entraîner le modèle, des librairies Python seront nécessaires, parmi ces dernières on peut compter: **Pytorch**, **Numpy**, **Pandas**, **Matplotlib** etc.

III - Planning prévisionnel



IV - Conclusion

Résultats attendus

Pour statuer sur la qualité du modèle on dispose de plusieurs indicateurs:

Précision élevée : l'objectif principal est de pouvoir identifier correctement la langue dans un enregistrement audio avec une précision élevée.

Robustesse : aux variations de la voix : il est donc important que le modèle soit capable de reconnaître les différents accents et dialectes.

Scalabilité : le modèle doit être capable de gérer de nouvelles langues et de nouveaux accents avec des données supplémentaires.

Temps de réponse rapide : pour des applications en temps réel, il est important que le modèle soit capable de traiter les enregistrements audio et de fournir des résultats rapidement.

Capacité à fonctionner avec des enregistrements de qualité variable : il est important que le modèle soit capable de fonctionner avec des enregistrements de qualités variables.

La précision est l'indicateur central de la réussite du projet mais les autres indicateurs sont les témoins de la qualité de ce dernier, il faudra alors les prendre en compte.

Limitations et améliorations possibles

Le projet est ambitieux et peut présenter certains défauts pouvant freiner sa réussite. Pour cela il est nécessaire d'y réfléchir en amont et de trouver des solutions adéquates.

1. Augmenter la taille du dataset
2. Comparer différents types d'architectures ou les combiner.
3. Trouver et implémenter un autre optimisateur de paramètres et une fonction d'erreurs différents.
4. Effectuer l'ajustement des hyperparamètres.

À terme, quand le modèle sera validé, l'objectif est de mettre à disposition cet outil en ligne sous forme d'upload de fichiers ou de capture audio instantanée. Il faudra alors réaliser un site sous Flask, un framework python, et trouver un moyen de déployer le modèle entraîné.

🔗Pour me contacter