

PLATEFORME DE PREDICTION ET DE RECONNAISSANCE DES IDENTITÉS

1. OBJECTIFS

Cette plateforme, basée sur de l'apprentissage supervisé, vise à permettre à déterminer automatiquement l'identité des personnes (Par exemple pays d'origine)

2. MODULES ET FONCTIONNALITÉS

M1. GESTION DES UTILISATEURS

Ce module gère l'inscription et l'authentification des utilisateurs sur la plateforme.

F1.1. INSCRIPTION

- Le visiteur peut s'inscrire sur la plateforme de deux manières :
 - en mentionnant son adresse e-mail, son mot de passe et son numéro de téléphone. Le compte n'est créé que lorsque l'utilisateur valide son adresse e-mail (mail de validation).
 - via un compte tiers de confiance : Facebook ou Gmail

F1.2. CONNEXION

- Le visiteur se connecte à son compte via son email et mot de passe.
- Il peut optionnellement cocher une case "rester connecté" pour garder sa session active.
- Il est à noter que si l'utilisateur se connecte d'un PC / tablette / Téléphone n'ont connu, un message SMS est envoyé au numéro renseigné lors de l'inscription pour l'avertir d'une connexion sur un terminal non connu

F1.3. RÉINITIALISATION DU MOT DE PASSE

En cas d'oubli de mot de passe, le visiteur peut demander à réinitialiser son mot de passe en fournissant son adresse e-mail. Un email sera alors envoyé pour lui permettre de changer le mot de passe.

F1.4. ESPACE PROFIL

Chaque joueur dispose de sa propre page profil qui contient ses informations personnelles (email, téléphone) et l'historique des connexions et les devices associés

M2. GESTION DU DATASET DE DONNÉES

- l'utilisateur peut consulter et enrichir le dataset de données à disposition

F2.1. CONSULTATION DU DATASET DE DONNÉES

L'utilisateur peut consulter le dataset des données sous forme de tableau. L'affichage du tableau peut se faire de deux manières :

- tableau est paginé avec un choix du nombre de résultats à afficher par page .
- Infinite scroll

Le tableau affiché contient les informations suivantes :

- Prénom
- Nom
- Date de naissance + age entre parenthèse
- Ville
- Province/Région
- Pays
- Nombre d'années d'expérience (optionnel)

F2.2. RECHERCHE AU NIVEAU DU DATASET DE DONNÉES

Le champs de recherche se base sur l'ensemble des champs du tableau (implémentation d'Elastic Search)

La recherche se base sur de l'autocomplete

F2.3. ENRICHISSEMENT DU DATASET

Il est possible d'étoffer le dataset d'apprentissage de deux manières :

- Upload d'un fichier Excel avec en colonne les informations suivantes : prénom, nom (soit dans une seule colonne ou en deux colonnes), date de naissance, ville, province/région, pays
- Upload en lot de CVs (Word et PDF)

F2.4. TRAITEMENT OCR POUR LES CVs PDF

Les CVs PDF uploadés auront un traitement OCR afin de pouvoir extraire les mots à stocker au niveau de la base de données

F2.5. EXTRACTION DES DONNÉES PERTINENTES DES DOCUMENTS UPLOADÉS

Les informations qui sont extraites des documents uploadés sont :

- Prénom
- Nom
- Date de naissance + age entre parenthèse
- Ville
- Province/Région
- Pays
- Nombre d'années d'expérience (calculé selon le parcours sur le CV)

M3. APPRENTISSAGE SUPERVISÉ

Ce module vise à utiliser ou à implémenter un algorithme d'apprentissage supervisé sur l'origine des identités de personnes en fonction de leur prénom et de leur nom.

L'objectif est que prédire l'origine de la personne et son âge en fonction des informations de prénom/nom

F3.1. ALGORITHME D'APPRENTISSAGE SUPERVISÉ

- En fonction du dataset fourni et qui est à étoffer, une étude et une implémentation de l'algorithme supervisé seront réalisées.

F3.2. VÉRIFICATION DE L'IDENTITÉ

- Cette fonctionnalité vise à permettre à l'utilisateur de saisir un prénom et un nom pour déterminer l'origine
- En se basant sur l'algorithme, le résultat retourné sera une prédiction de la région, du pays et de l'âge de la personne

- En fonction du résultat retourné et si nous connaissons les résultats attendus, nous pouvons corriger les valeurs retournées pour enrichir l'apprentissage de l'algorithme.