

CERTIFICATION DÉVELOPPEUR EN INTELLIGENCE ARTIFICIELLE ET DATA SCIENCE RNCP 36581

BLOC DE COMPÉTENCES E6.2 : Développer un modèle prédictif d'une solution I.A

Cahier des Charges de la MSPR « Développement d'un modèle d'apprentissage répondant au besoin d'une solution IA A partir des éléments de contextes fournis (Cahier des charges données d'entrée) »

COMPÉTENCES ÉVALUÉES :

- Générer des données d'entrée, récolter et adapter les types de données traitées nécessaires au modèle
- d'apprentissage en utilisant des approches et des outils adaptés
- Paramétrer un environnement de codage (Framework) adéquat pour développer le modèle d'apprentissage
- Coder le modèle d'apprentissage choisi (réseau de neurones, boosting, forêt aléatoire, arbre de décision, etc) en maîtrisant les différentes architectures dans un environnement de développement.
- Réaliser et paramétrer une procédure d'entraînement adéquate d'un modèle d'apprentissage : supervisé, semi-supervisé, non supervisé, en sélectionnant des données d'apprentissages les plus adéquats aux besoins d'analyse.
- Réaliser une phase de test en choisissant une méthode appropriée afin d'analyser la performance du modèle de données.
- Ajuster l'apprentissage du modèle (entraînement) à partir du taux d'apprentissage et des résultats obtenus.

PHASE 1 : PRÉPARATION DE CETTE MISE EN SITUATION PROFESSIONNELLE RECONSTITUÉE

- Durée de préparation :
 - 31 heures
- Mise en oeuvre :
 - Travail d'équipe constituée de 4 apprenants-candidats (5 maximum si groupe impair)
- Résultat attendu :
 - Data augmentation et rééquilibrage des labels si nécessaire
 - Choix des métriques d'évaluation
 - Choix du modèle le plus performant, et optimisation
 - Rédaction d'un court compte-rendu sur les performances du modèle

PHASE 2 : PRÉSENTATION ORALE COLLECTIVE + ENTRETIEN COLLECTIF

- **Durée totale par groupe** : 30 mn se décomposant comme suit :
 - 10 mn de soutenance orale par l'équipe.
 - 20 mn d'entretien collectif avec le jury (questionnement complémentaire).
 - Objectif : mettre en avant et démontrer que les compétences visées par ce bloc sont bien acquises.
- **Jury d'évaluation** : 2 personnes (binôme d'évaluateurs) par jury – Ces évaluateurs ne sont pas intervenus durant la période de formation et ne connaissent pas les apprenants à évaluer.

I - PRÉSENTATION DE L'ENTREPRISE ET DE SON ACTIVITÉ

Préambule : L'entreprise choisie pour cette MSPR est fictive, les prénoms sont fictifs, toute ressemblance à un cas réel serait purement fortuite.



WildLens est une association française engagée dans la protection animale dans les régions sauvages. Elle collecte des fonds pour financer ses actions et mène des campagnes de sensibilisation en forêt pour informer le public sur les enjeux de la conservation de la faune sauvage.

L'association souhaite tirer parti des nouvelles technologies pour sensibiliser davantage le public sur la nature qui nous entoure. WildLens veut développer une application d'identification des traces de pas pour sensibiliser le public à la préservation de la faune sauvage de façon ludique, en leur montrant les empreintes laissées par ces animaux dans leur habitat naturel. Chaque utilisateur pourra ainsi scanner les empreintes qu'il croise, afin de connaître l'animal qui l'a laissée et accéder à quelques informations intéressantes.

Cette application permettra en outre de recueillir des données précises sur les animaux, telles que leur fréquence de passage et leur emplacement, qui pourraient être utiles pour suivre leur évolution et établir des plans de préservation efficaces.

WildLens fait appel à votre expertise pour le développement de ce projet

II- SPÉCIFICATIONS DE L'APPLICATION – LES FONCTIONNALITÉS

L'application doit permettre l'identification d'empreintes sauvages, ainsi que le recueil de données pertinentes pour les équipes de WildLens.

La page principale doit permettre à l'utilisateur de prendre une photo d'empreinte. Une fois une empreinte prise en photo, la page affichera le nom de l'animal détecté ainsi que les données pertinentes associées. La maquette ci-dessous a été réalisée à titre d'exemple.

A chaque identification d'une empreinte, les informations relatives à la prise de photo (date, heure, localisation) ainsi que les résultats (animal identifié) doivent être stockées dans une base de données accessible par WildLens.



Design

Le design général de l'application est laissé libre à l'équipe en charge du développement. Il doit utiliser à minima le logo et les couleurs de WildLens. Différentes déclinaisons du logo sont disponibles dans le dossier Design des ressources fournies.

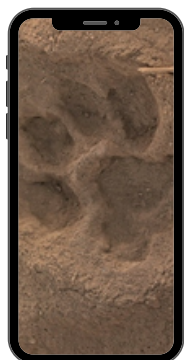
Plateforme

A terme, l'application sera disponible sur Android et iOS. Cependant, dans le cadre de la mise en œuvre du modèle de détection d'empreintes, la direction souhaite **seulement réaliser une version web qui s'adapte aux petits écrans**.

Spécifications du modèle de reconnaissance – Le Modèle

A partir d'une photo d'empreinte laissée dans la nature, le modèle devra être capable d'identifier l'animal qui l'a laissée. Il sera demandé aux utilisateurs de l'application de prendre en photo uniquement l'empreinte, et non l'ensemble des traces (Exemple ci-dessous). L'équipe de développement devra le prendre en compte lors de son travail sur le jeu de données.

Tous les choix relatifs à la construction du modèle (méthode, outils, etc.) seront laissés à l'équipe en charge de son développement.



✓
Exemple de photo pouvant
être prise par les utilisateurs



✗
Exemple de photo
non valide

Données Identification

Pour vous aider dans cette tâche, les experts animaliers de chez WildLens vous fournissent un jeu de données de photos labellisées d'empreintes animales. Vous les trouverez dans le sous-dossier Mammifères du dossier de ressources fourni.

Les données fournies comportent 252 photos d'empreintes pour 13 espèces de mammifères (Castor, Chat, Chien, Coyote, ..., Rat, Raton Laveur, Renard). L'application se contentera d'identifier ces 13 espèces les plus communes dans un premier temps.

Les choix de nettoyage des données et de data augmentation seront laissés à l'équipe en charge du développement. Notez cependant que le jeu de données peut contenir des photos avec des éléments extérieurs (watermark, objets, etc.) ou ne correspondant pas à des photos d'une seule empreinte.

- Source des photos : <https://naturetracking.com/> et <https://www.inaturalist.org/observations>
- Données complémentaires : Le fichier `infos_especes.csv` du dossier de ressources contient les informations à afficher pour chaque espèce.

Le projet a été découpé en trois lots qui seront livrés au fur et à mesure, permettant à WildLens de suivre l'évolution du projet.

💡 Lien vers le dossier de ressources :
<https://www.dropbox.com/sh/su3xv1drslb09qj/AAADjEysF3vr7asXMUc6ExwQa?dl=0>

III- EXPRESSION DE LA DEMANDE

Le lot numéro 2 a pour but de mettre en place le modèle de catégorisation le plus performant.

IV-ANALYSE DU PROBLÈME

L'équipe de développement devra s'assurer que les données aient été correctement préparées et sont accessibles dans l'environnement de développement du modèle.

Si cela semble pertinent, il est recommandé de mettre en œuvre des techniques de data augmentation afin de disposer d'un jeu de données plus riche. Des techniques permettant un rééquilibrage des labels peuvent aussi être mises en place.

V- MODÈLE

Il est important de définir une métrique, la plus pertinente selon la problématique, afin de mesurer les performances d'un modèle. Une fois la métrique définie, l'équipe pourra tester différents modèles de classification pour déterminer le plus performant.

Pré-traitements

Différents pré-traitements peuvent être mis en place sur les images (effets de flou pour atténuer la texture du sol, effets de contraste pour faire ressortir l'empreinte, etc.). Il est recommandé de tester différentes techniques pour comparer les performances obtenues.

VI- PERFORMANCES

- **FINE-TUNING**

- Une fois que l'algorithme le plus optimal a été trouvé, il est recommandé d'avoir recours à des méthodes de recherches des hyperparamètres optimaux (grid search, par exemple). Cela peut permettre d'améliorer une dernière fois les performances du modèle.

- **COMPTE-RENDU**

- L'équipe fera enfin un court compte-rendu à destination de WildLens présentant les performances du modèle choisi, ainsi que leur signification. Ce compte-rendu s'adressera à des personnes familières au monde de la data science mais non expertes

