

Rapport de Projet : Workflow de Gestion des Tickets de Support

Projet de Classification des Manchots avec Machine Learning

1. Introduction

Dans ce projet, nous avons développé une application interactive de classification d'espèces de manchots en utilisant l'algorithme Random Forest. L'application est implémentée avec Python et Streamlit pour offrir une interface intuitive permettant aux utilisateurs d'entrer les caractéristiques morphologiques d'un manchot et d'obtenir une prédiction de son espèce.

2. Objectifs du Projet

- *Construire un modèle de Machine Learning supervisé pour prédire l'espèce d'un manchot.*
- *Créer une interface interactive permettant d'entrer des caractéristiques et d'afficher la prédiction.*
- *Visualiser les données et les résultats de la classification de manière claire.*

3. Données Utilisées

Nous avons utilisé le dataset "Penguins" provenant de [Kaggle](#). Ce dataset contient 344 observations et les caractéristiques suivantes :

- *Caractéristiques numériques : longueur et profondeur du bec (mm), longueur des nageoires (mm), masse corporelle (g).*
- *Caractéristiques catégoriques : île d'origine et sexe du manchot.*
- *Variable cible : espèce du manchot (Adélie, Chinstrap, Gentoo).*

4. Prétraitement des Données

Avant d'entraîner le modèle, nous avons effectué les transformations suivantes :

- *Encodage des variables catégoriques (island et sex) en variables indicatrices avec `pd.get_dummies()`.*
- *Encodage de la variable cible (species) en valeurs numériques :*
 - *Adélie → 0*
 - *Chinstrap → 1*
 - *Gentoo → 2*
- *Normalisation des données (optionnelle pour améliorer la performance du modèle).*

5. Modèle de Machine Learning

Nous avons choisi un Random Forest Classifier pour son efficacité en classification et sa robustesse face aux données bruitées. Le modèle a été entraîné sur les caractéristiques prétraitées :

```
from sklearn.ensemble import RandomForestClassifier
```

```
clf = RandomForestClassifier()
```

```
clf.fit(X, y)
```

6. Interface Utilisateur avec Streamlit

L'application Streamlit permet aux utilisateurs d'entrer les caractéristiques d'un manchot via des widgets interactifs (sliders, menus déroulants). Une fois les valeurs saisies, le modèle prédit l'espèce du manchot et affiche les probabilités associées.

Exemple de widgets :

```
bill_length_mm = st.slider('Bill length (mm)', 32.1, 59.6, 43.9)
```

```
body_mass_g = st.slider('Body mass (g)', 2700.0, 6300.0, 4207.0)
```

L'affichage des résultats inclut :

- *Probabilités de chaque espèce sous forme de tableau.*
- *Affichage du nom de l'espèce prédite avec st.success().*

7. Résultats et Analyse

Le modèle fournit des prédictions précises avec un bon niveau de confiance, et les probabilités associées permettent de visualiser les incertitudes du modèle.

8. Améliorations Possibles

- *Optimisation des hyperparamètres (n_estimators, max_depth) pour améliorer la performance.*
- *Standardisation des caractéristiques numériques.*
- *Ajout d'autres modèles de classification (SVM, KNN) pour comparer les performances.*
- *Évaluation avancée : matrice de confusion, scores de précision, rappel et F1-score.*

9. Conclusion

Ce projet démontre l'application pratique du Machine Learning à la classification d'espèces de manchots, avec une approche interactive grâce à Streamlit. Il met en avant les compétences en data science, modélisation et développement d'interfaces utilisateurs pour des applications de Machine Learning.