

Projet Maching Learning

Master en Science de données et système intelligent

Ballet Floral : Un Pas Vers la Détection Délicate des Types de Fleurs

Réalisé par:

- SALWA KBIRI ALAOUI
- RABAB KAF
- CHAYMAE ISMAILI
- ILHAME JABBARI

Encadré par:

- Mr. JAMAL KHARROUBI

Outils d'extraction	Modèles d'entraînement	Temps d'extraction	Temps d'entraînement	Accuracy	members	materiels utilisé
scikit-image	SVM	22 min	30 min	52%	Salwa Kbiri Alaoui 100%	DELL xps RAM :16 Go CPU : 475 Go
scikit-image	RandomForest	21 min	25 min	54%	Ilhame Jebbari :100%	Dell RAM :8 Go CPU : 256 Go
OpenCV	RandomForest	2 min	3 min	62%	Rabab Kaf :100%	Assus VivoBook RAM :16 Go CPU :337 Go
Architecture de Réseau CNN	RandomForest	40 min	35 min	64%	Chaimae Ismaili :100%	Lenovo. RAM 16GB Disque 952 Go
VGG	RandomForest	45 min	50 min	73%	Chaimae Ismaili :100%	Lenovo. RAM 16GB Disque 952 Go
ResNet50	KNN	47 min	180 min	86%	Salwa kbiri Alaoui :100%	DELL xps RAM :16 Go CPU : 475 Go
VGG	KNN	40 min	17 min	86%	Rabab Kaf :100%	Assus VivoBook RAM :16 Go CPU :337 Go
ResNet50	RandomForest	50 min	161 min	87%	Ilhame Jebbari :100%	Dell RAM :8 Go CPU : 256 Go
ResNet50	SVM	47min	161min	93%	Ilhame Jebbari :40% Rabab Kaf :20% Chaime Ismaili :20% Salwa Kbiri Alaoui :20%	Dell RAM :8 Go CPU : 256 Go

Résumé

Ce rapport présente les différentes étapes de réalisation d'un projet de machine learning dans le cadre de notre formation en Master Science de Données et Systèmes Intelligents.

Notre projet vise le développement d'une application web reposant sur un modèle de machine learning capable d'identifier et de classer différents types de fleurs, en se basant uniquement sur des caractéristiques visuelles. Nous avons suivi un processus débutant par la collecte et le nettoyage des données, puis l'extraction des caractéristiques et l'entraînement de plusieurs modèles pour obtenir un résultat satisfaisant.

Notre approche s'appuie sur l'utilisation de modèles de deep learning et de machine learning, ce qui nous a permis d'obtenir des résultats prometteurs dans la reconnaissance des fleurs.

Table des matières

Introduction générale	1
Chapitre 1 : Contexte général du projet	2
Objectif	2
1.1 Jeu de données.....	2
2.3 Optimisation : Choix de la Meilleure Combinaison.	3
Chapitre 3 : Réalisation de l'application	5
3.1 Environnement de réalisation	5
3.2 Langages et technologies utilisées	5
3.3 Présentation de l'application	6
Conclusion générale	7

Introduction générale

Dans le cadre de notre formation universitaire en Master en génie informatique à la faculté des sciences et techniques de Fès, nous nous sommes lancés dans un projet de recherche ambitieux : appliquer les techniques de Machine Learning à la reconnaissance et à la classement des différentes espèces de fleurs à partir d'images. Ce projet représente une étape cruciale de notre parcours académique, stimulé par notre passion pour l'innovation technologique et notre désir de contribuer au progrès scientifique dans le domaine de l'intelligence artificielle appliquée à la botanique.

Notre première étape fut d'analyser le contexte et de cerner la problématique. Comment pouvons-nous développer un modèle de Machine Learning capable de distinguer avec une précision remarquable les différentes espèces de fleurs à partir d'images, offrant ainsi un outil précieux aux botanistes, aux jardiniers et aux amateurs de plantes? Nous avons ensuite parcouru un processus intensif de tests et de comparaisons entre différents modèles, évaluant leur performance et leur adaptabilité à notre problème spécifique.

Notre choix de modèles s'est appuyé sur des critères rigoureux, tenant compte à la fois de leur précision de classement, de leur capacité à gérer des données d'entrée complexes telles que des images, et de leur potentiel d'adaptation à des contextes variés. Nous avons alors affiné notre sélection pour retenir les modèles les plus performants et les mieux adaptés à notre projet.

Enfin, nous avons mis en œuvre ces modèles choisis, réalisant des évaluations approfondies de leur efficacité, de leur robustesse et de leur capacité à générer des résultats précis. Ces résultats ont été présentés de manière détaillée, accompagnés de remarques pertinentes sur les données d'entrée utilisées, les limites rencontrées et les perspectives d'amélioration pour des travaux futurs.

Ce projet de détection des types de fleurs par le biais du Machine Learning est le fruit d'un travail méthodique, passionné et rigoureux. Il démontre notre capacité à relever des défis complexes en intelligence artificielle, tout en mettant en lumière notre engagement envers l'innovation technologique et notre contribution au développement de solutions pratiques pour des domaines aussi variés que la botanique, l'agriculture et la préservation de l'environnement.

Chapitre 1 : Protocol expérimental

1 Objectif

L'objectif principal de cette application est de développer un système automatisé capable de reconnaître et de classer différentes espèces de fleurs à partir d'images, tout en fournissant des informations détaillées sur chaque espèce. Cette solution vise à faciliter l'identification et la compréhension des fleurs pour un large éventail de personnes, notamment les botanistes, les jardiniers et les passionnés de plantes. En combinant des techniques avancées d'apprentissage automatique avec une analyse approfondie des données, l'objectif est de fournir une assistance précieuse pour surmonter les défis liés à l'identification et à la connaissance des différentes espèces florales.

1. Jeu de données :

Notre jeu de données comprend un total de 5 534 images de fleurs réparties en 11 classes différentes. Cinq des classes ont été téléchargées à partir du site Kaggle, tandis que les six autres classes ont été collectées manuellement sur Internet. Les cinq classes téléchargées depuis Kaggle sont **Daisy**, **Tulip**, **Rose**, **Sunflower** et **Dandelion**, avec environ 600 images par classe. Les six autres classes collectées manuellement comprennent des fleurs telles que **Amirillis**, **Forget me nots**, **Jasmin**, **Lotus**, **Osteospermum** et **Pansy**, avec un nombre variable d'images par classe.

Avant d'inclure les images dans le jeu de données, une étape de nettoyage a été effectuée pour éliminer les mauvaises images.

Les images ont des résolutions variées et des proportions différentes, allant de 320x240 pixels pour les images les plus petites à des résolutions plus élevées pour certaines images téléchargées.

Ce jeu de données composite offre une diversité d'images de fleurs provenant de différentes sources, ce qui en fait une ressource précieuse pour l'entraînement et l'évaluation de modèles de classement des fleurs.

Le jeu de données initial téléchargé à partir de Kaggle était divisé en deux parties : entraînement et test. Cependant, nous avons choisi de regrouper toutes les images dans un seul dossier pour chaque classe, tant pour les données téléchargées que pour celles collectées manuellement.

Pour diviser le jeu de données en ensembles d'entraînement et de test, nous avons appliqué la règle de division 1/3 et 2/3. Cela signifie que pour chaque classe, environ un tiers des images ont été réservées pour les données de test et les deux tiers restants ont été utilisés pour les données d'entraînement. Cette approche nous permet d'avoir une répartition équilibrée des données pour l'entraînement et le test de nos modèles de classement des fleurs.

2.1 Optimisation : Choix de la Meilleure Combinaison : SVM et ResNet50 avec Validation Croisée

Après avoir exploré plusieurs combinaisons d'outils d'extraction de caractéristiques et de modèles d'entraînement pour la détection de fleurs, nous avons sélectionné une approche prometteuse combinant ResNet50 et SVM avec validation croisée (cross-validation).

Explorons comment cette combinaison a été implémentée dans notre code :

Tout d'abord, nous avons utilisé le modèle pré-entraîné ResNet50 pour extraire les caractéristiques des images de fleurs. ResNet50 est un réseau de neurones convolutif profond qui a démontré d'excellentes performances dans la classement d'images. En utilisant ce modèle, nous avons pu capturer des représentations riches et informatives des images de fleurs, qui serviront de données d'entrée pour notre modèle d'apprentissage automatique.

Ensuite, nous avons choisi le modèle de Support Vector Machine (SVM) comme algorithme de classement. Pour améliorer les performances de ce modèle, nous avons utilisé la technique de validation croisée pour rechercher les meilleurs hyperparamètres. Nous avons utilisé la classe GridSearchCV de la bibliothèque scikit-learn, qui nous a permis d'explorer différentes combinaisons de valeurs pour les paramètres C et gamma du SVM tout en utilisant la validation croisée pour évaluer la performance du modèle.

Après avoir effectué cette recherche exhaustive, nous avons identifié les meilleurs paramètres qui maximisent la précision du modèle sur les données d'entraînement. Une fois que nous avons obtenu le meilleur modèle SVM avec les hyperparamètres optimisés grâce à la validation croisée, nous l'avons évalué sur un ensemble de test séparé pour estimer sa performance sur des données non vues.

La précision du modèle sur l'ensemble de test était de **0.96**, ce qui représente un score très élevé et prometteur. Enfin, pour évaluer plus en détail les performances du modèle, nous avons affiché une matrice de confusion qui nous permet de visualiser les prédictions du modèle par rapport aux vraies étiquettes des fleurs.

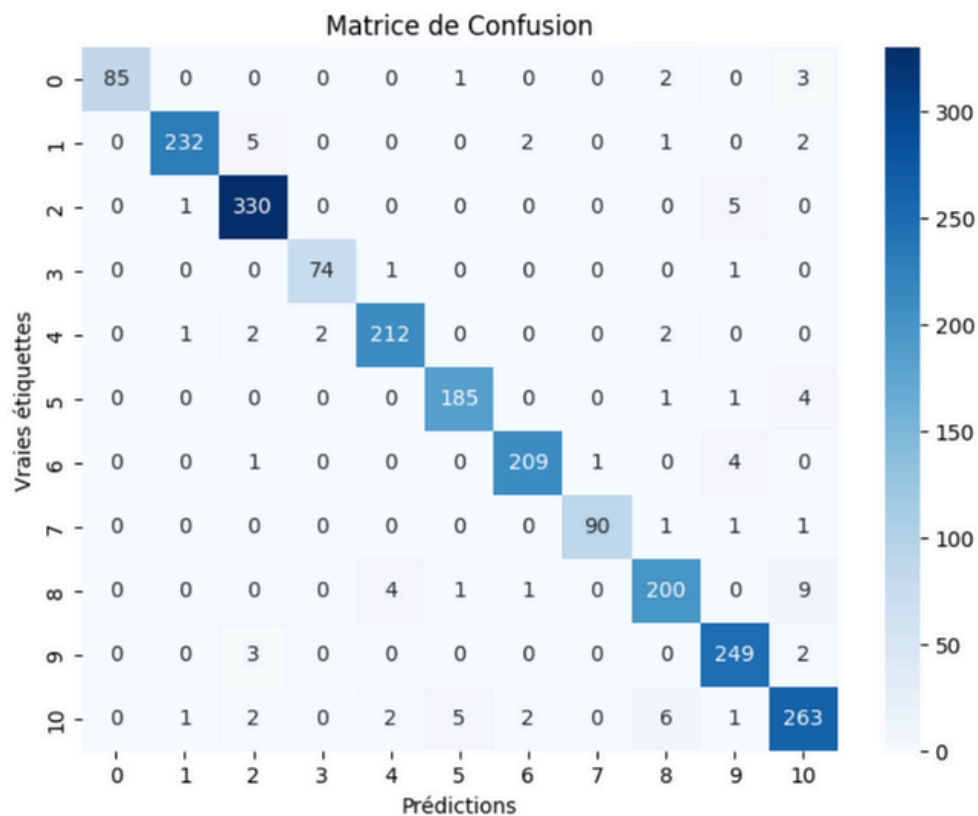


Figure 1: Matrice de Confusion

nous pouvons interpréter la matrice de confusion comme suit :
par exemple pour la classe 1:

Le modèle a correctement prédit 232 images appartenant à la classe 1. Cependant, il a également confondu 5 images de la classe 2 avec la classe 1, ainsi que 2 images de la classe 6 et 2 images de la classe 10. En général, en examinant les performances pour chaque classe, nous constatons que notre modèle a une précision globale de 95% dans la détection des fleurs, ce qui indique sa fiabilité dans la reconnaissance des différentes classes de fleurs.

Cette analyse approfondie confirme la robustesse et l'efficacité de la combinaison choisie de ResNet50 et SVM pour la détection de fleurs par Machine Learning, et met en évidence l'importance de la validation croisée dans l'optimisation des hyperparamètres du modèle.

Chapitre 3 : Réalisation de l'application

1 Environnement de réalisation

VSCode:

L'environnement de réalisation de Visual Studio Code offre une interface personnalisable, des extensions polyvalentes et un éditeur de texte puissant avec intégration Git, facilitant ainsi le développement efficace d'applications.



Figure 2: Logo VSCode

2 Langages et technologies utilisées

Python :

Python est largement utilisé en machine learning pour sa simplicité, sa syntaxe claire et sa richesse en bibliothèques. Grâce à des frameworks tels que TensorFlow, PyTorch et scikit-learn, Python permet de mettre en œuvre facilement des algorithmes d'apprentissage automatique pour la classification, la régression, le clustering et bien d'autres tâches, contribuant ainsi au développement d'applications innovantes dans des domaines tels que la reconnaissance d'images, le traitement du langage naturel et la recommandation personnalisée.

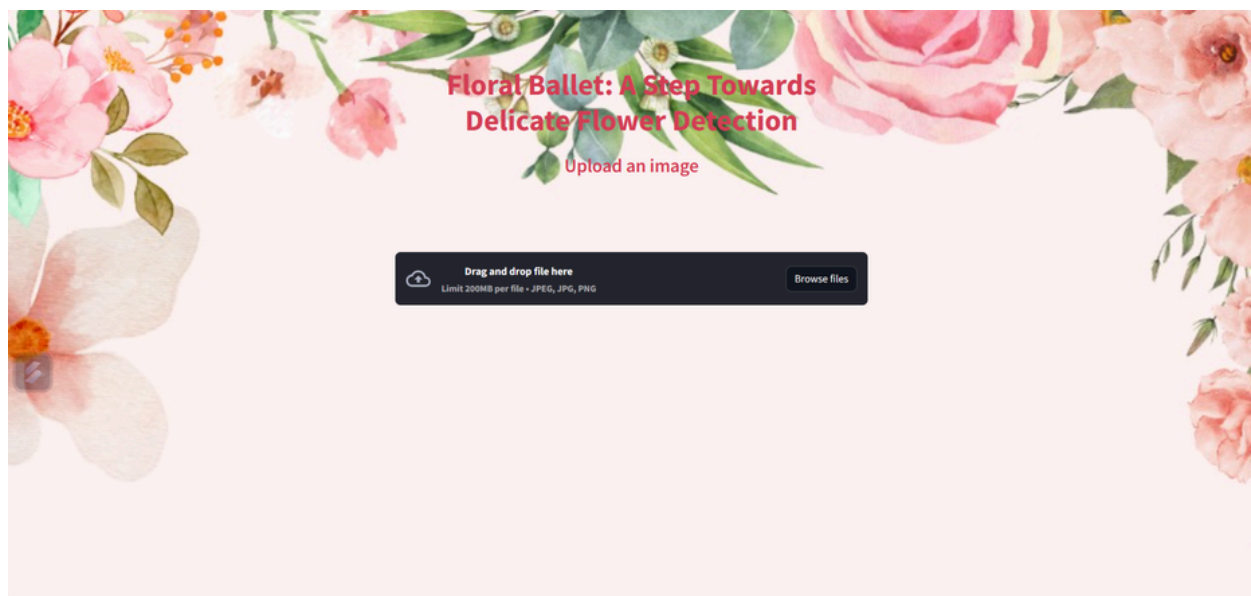


Figure 3: Logo Python

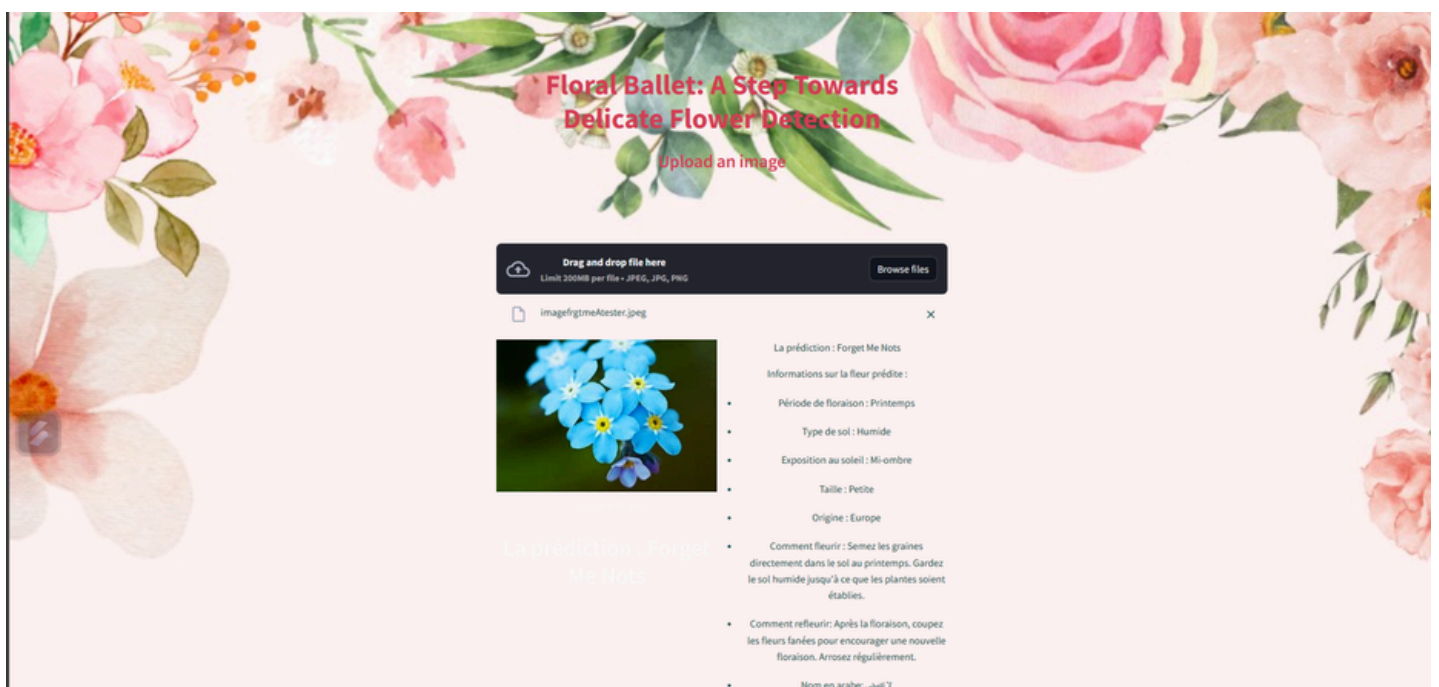
3 Présentation de l'application

La page d'accueil :

Cette page permet à l'utilisateur de télécharger une image d'une fleur afin de demander au modèle de prédire son type. Ensuite, elle affiche quelques informations sur ce type de fleur.



Voici un test avec l'image "Forget Me Nots". Le modèle parvient très bien à prédire le type de la fleur, et en plus, il affiche quelques informations utiles et intéressantes sur cette fleur.



Conclusion

En conclusion, ce projet de machine learning dans le cadre de notre formation en Master Science de Données et Systèmes Intelligents a été une expérience enrichissante et instructive. Notre objectif était de développer une application web capable d'identifier et de classer différents types de fleurs uniquement à partir de caractéristiques visuelles, en utilisant des techniques de deep learning et de machine learning.

À travers les différentes étapes de ce projet, de la collecte des données au déploiement de l'application web, nous avons acquis une compréhension approfondie des défis et des opportunités liés à la mise en œuvre de solutions de machine learning dans un contexte pratique. Nous avons exploré diverses techniques de prétraitement des données, d'extraction de caractéristiques et d'entraînement de modèles, en nous efforçant de trouver l'équilibre optimal entre la complexité du modèle et sa capacité à généraliser correctement.

Nos résultats ont été prometteurs, démontrant la capacité de notre modèle à reconnaître et à classer efficacement les différentes espèces de fleurs. Cependant, il reste encore des possibilités d'amélioration, notamment en explorant des architectures de réseaux neuronaux plus avancées, en augmentant la taille et la diversité du jeu de données, et en affinant les paramètres des modèles pour une performance optimale.

En fin de compte, ce projet a confirmé l'importance cruciale de chaque étape du processus de machine learning, depuis la collecte et le nettoyage des données jusqu'à l'évaluation et le déploiement des modèles. Il souligne également l'importance de la collaboration interdisciplinaire, combinant des compétences en science des données, en informatique et en domaines d'application spécifiques pour relever avec succès les défis complexes de l'analyse et de la compréhension des données.

En tant qu'équipe, nous sommes fiers du travail accompli et confiants dans le potentiel de notre application pour contribuer à des domaines tels que la botanique, l'agriculture et la préservation de l'environnement. Nous espérons que ce rapport servira de base pour d'autres travaux futurs dans le domaine de la reconnaissance d'images et du machine learning, et nous sommes impatients de continuer à explorer et à innover dans ce domaine passionnant.