



Projet 8

Réalisez un Dashboard et assurez une veille technique

Samira MAHJOUR

03/09/2024

SOMMAIRE



Contexte



Présentation du tableau de bord



Présentation du modèle Efficient Net



Présentation du dataset



Préparation et modélisation des données



Conclusion

CONTEXTE

➤ **Problématique :**

Prêt à Dépenser doit répondre à une demande croissante de transparence de ses clients concernant les décisions d'octroi de crédit. Cela nécessite un outil qui permet d'expliquer ces décisions de manière simple et compréhensible pour un public non expert.

➤ **Mission :**

Dashboard interactif : Concevoir et déployer un Dashboard accessible via le Cloud, permettant aux chargés de relation client de visualiser le score de crédit et de fournir une explication claire lors des rendez-vous avec les clients.

Veille technique : Réaliser une étude sur les nouvelles technologies de modélisation de données texte et image, et les comparer à une approche classique pour s'assurer que l'entreprise reste à la pointe de l'innovation.

➤ **Objectifs :**

Dashboard : Développer un outil intuitif et conforme aux normes d'accessibilité pour visualiser les scores de crédit et permettre leur interprétation pour les clients.

Veille technique : Analyser une technologie récente de modélisation et tester son potentiel pour améliorer les performances des modèles actuels de l'entreprise.

PRESENTATION DU TABLEAU DE BORD

[https://bit.ly/m/Pret a depenser dashboard](https://bit.ly/m/Pret_a_depenser_dashboard)

Estimation du risque de non-remboursement

Entrez le SK_ID_CURR:

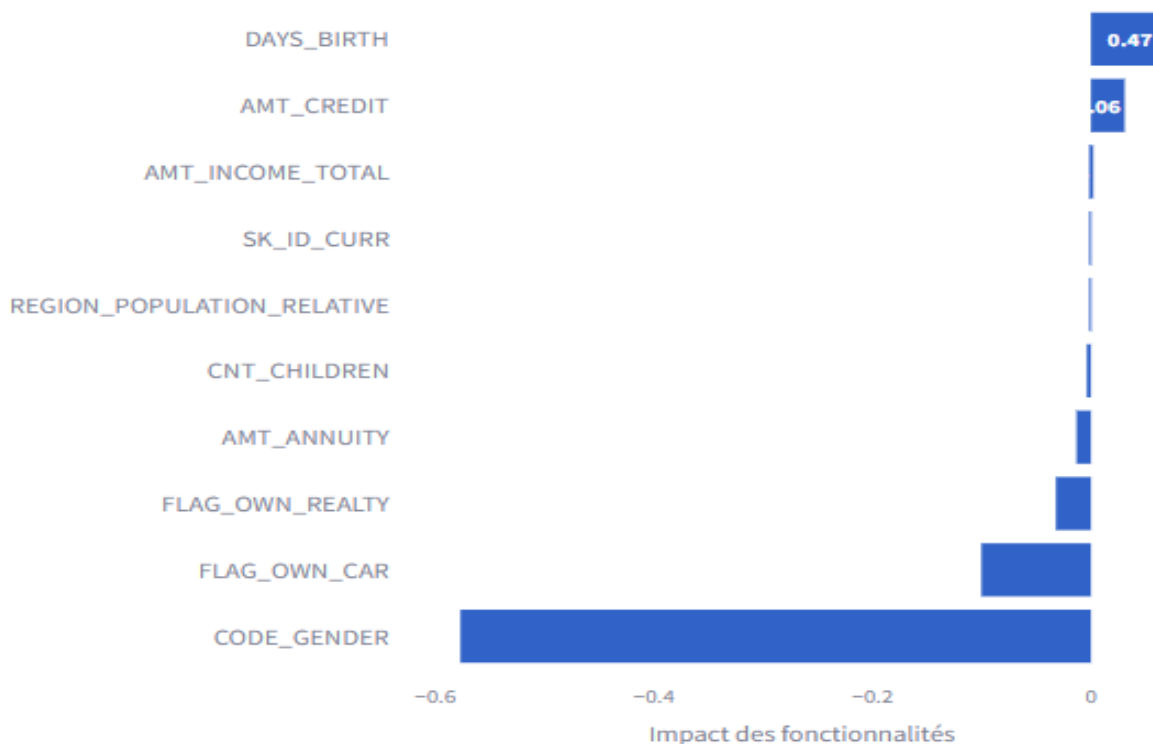
157083

Ru
n

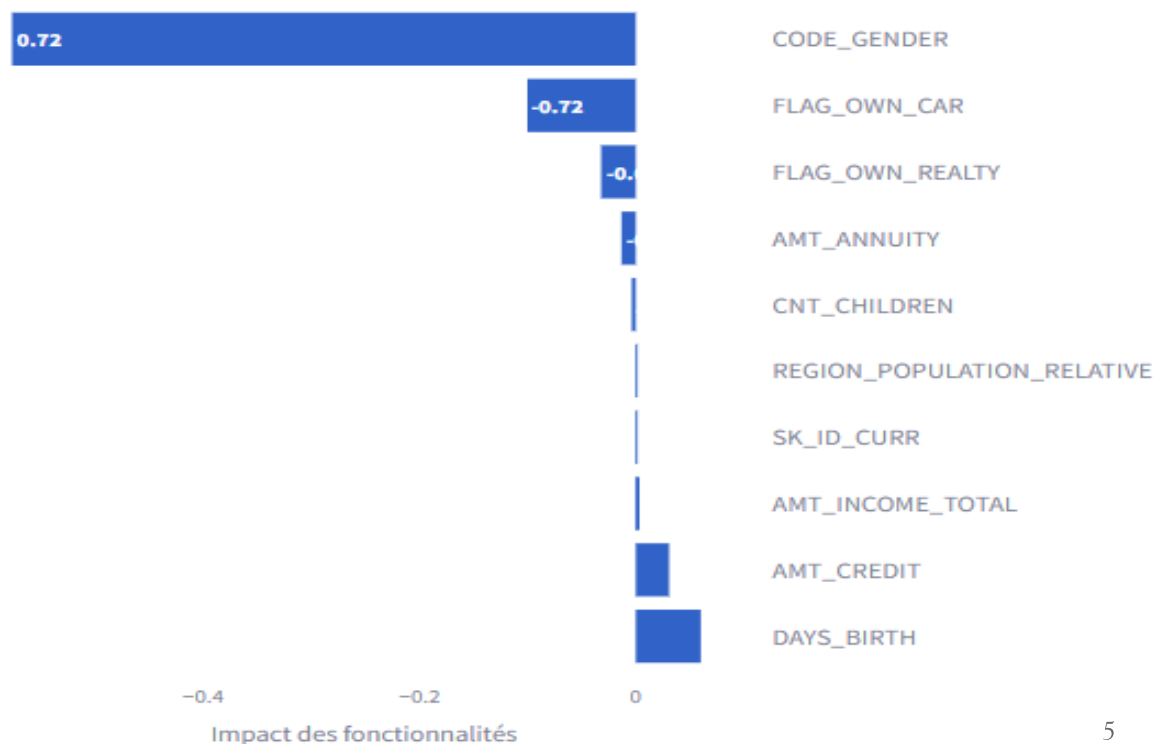
La probabilité que ce client ne puisse pas rembourser son crédit est de **0.09%** (tolérance max: 48%)

Le prêt sera accordé.

Top 10 des fonctionnalités augmentant le risque de non-remboursement



Top 10 des fonctionnalités réduisant le risque de non-remboursement



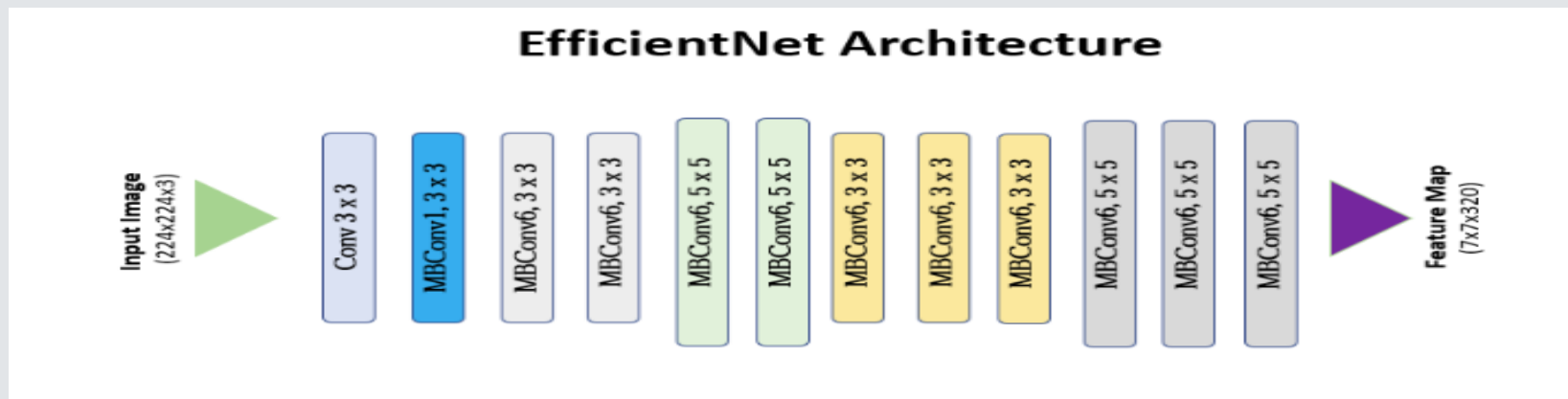
PRESENTATION DU MODELE Efficient Net

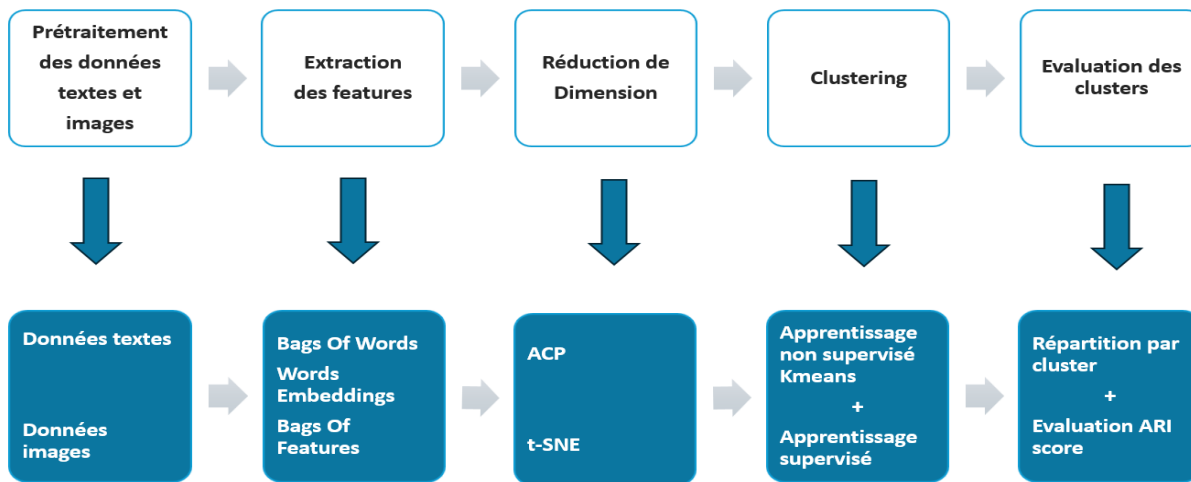
- Exploration d'une technique récente de modélisation : Efficient Net

- **EfficientNet** : Architecture de réseau de neurones qui ajuste automatiquement la taille du modèle en fonction des ressources pour maximiser la performance tout en réduisant la complexité et le temps d'entraînement .

- Développé par Google
- Conçu pour optimiser la taille et la performance des réseaux de neurones
- Utilise une approche de redimensionnement efficace pour ajuster la profondeur, la largeur, et la résolution du modèle
- Vise à obtenir de meilleures performances avec moins de ressources et un temps d'entraînement réduit
- Adapté pour des tâches de classification d'images à grande échelle

➤ Analyse, test et comparaison avec une approche classique utilisée précédemment






PRESENTATION DU DATASET

Projet 6 - Classifier
automatiquement des biens
de consommation

Présentation du dataset

Projet 6 - Classifier automatiquement des biens de consommation

Données issues de la base FlipKart  1 fichier csv + 1 dossier des images

- Un fichier CSV contenant 1050 articles
- Un dossier comprenant 1050 images
- Aucune donnée en double
- Aucune restriction liée à la propriété intellectuelle des données

Ce qui nous intéresse

- ✓ Données textuelles : nom, description
- ✓ Données visuelles : image des produits
- ✓ Les catégories des produits

PREPARATION ET MODELISATION DES DONNEES

APPROCHES DE TRAITEMENT D'IMAGES

```
graph TD; A[APPROCHES DE TRAITEMENT D'IMAGES] --> B[VGG16 avec intégration de data augmentation]; A --> C[EfficientNet-B7]; D[Normalisation des images] --- B; D --- C; E[Encodage des étiquettes de catégories] --- B; E --- C; F[Division en ensemble d'entraînement et de test] --- B; F --- C;
```

VGG16 avec intégration
de data augmentation

EfficientNet-B7

Normalisation
des images

Encodage des
étiquettes de
catégories

Division en ensemble
d'entraînement et de test

APPROCHE VGG16

Modèle pré-entraîné sur ImageNet :

Utilisé pour extraire les caractéristiques visuelles des images, basé sur un réseau de neurones convolutifs profond avec 16 couches.

Data Augmentation :

Techniques appliquées pour enrichir les données d'entraînement (rotation, zoom, renversement horizontal, translation) afin de réduire l'overfitting.

Fine-tuning :

Adaptation des couches supérieures du modèle à notre problème spécifique en gardant les premières couches gelées pour préserver les poids appris sur ImageNet.

Optimiseur Adam :

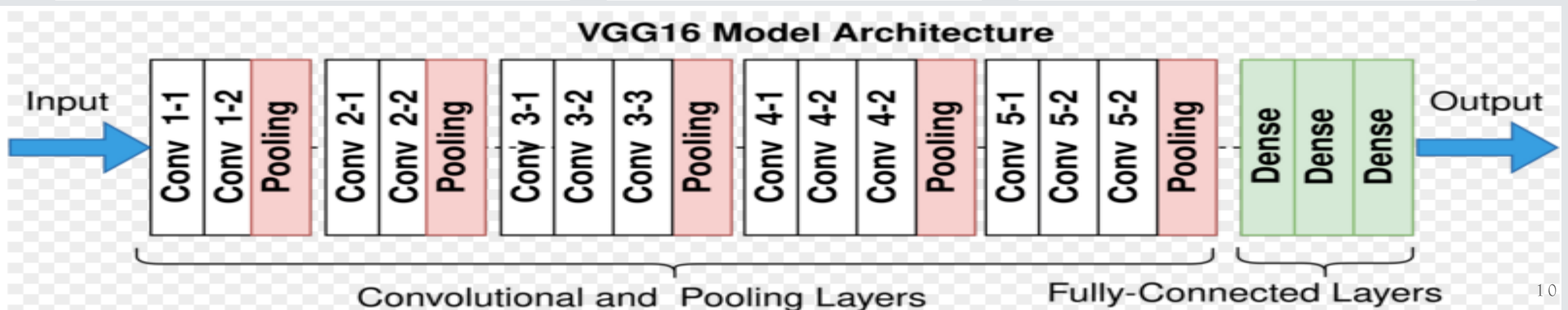
Méthode efficace pour ajuster les poids du modèle pendant l'entraînement.

Fonction de perte :

Categorical Cross-Entropy, adaptée pour la classification multi-classes.

Résultats :

Fournit une bonne précision sur des tâches de classification d'images complexes, bien que plus lent que les architectures récentes comme EfficientNet.



APPROCHE EfficientNet

Modèle pré-entraîné sur ImageNet :

EfficientNet utilise un modèle déjà entraîné sur un large ensemble d'images pour extraire des caractéristiques visuelles avec une grande efficacité.

Data Augmentation :

Techniques appliquées pour enrichir les données d'entraînement (rotation, zoom, translation, etc.) afin de rendre le modèle plus robuste et réduire le surapprentissage.

Mise à l'échelle équilibrée :

Agrandit la profondeur, la largeur et la résolution du modèle de manière proportionnelle pour optimiser les performances tout en réduisant le nombre de paramètres.

Optimiseur RMSprop :

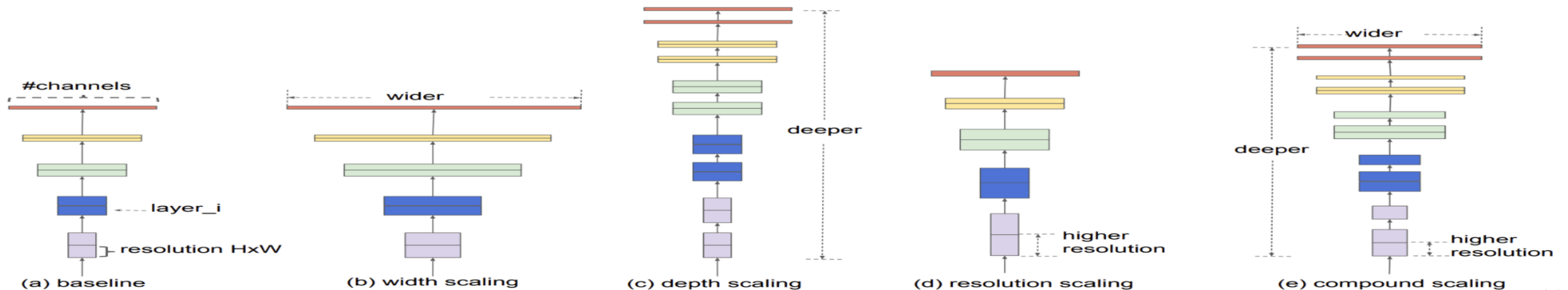
Utilisé pour ajuster les poids du modèle pendant l'entraînement, il est bien adapté aux réseaux de neurones profonds.

Fonction de perte :

Categorical Cross-Entropy, adaptée pour la classification multi-classes.

Résultats :

Très performant, avec une précision élevée et une meilleure efficacité que des modèles plus anciens comme VGG16.



METRIQUES D'EVALUATION ET RESULTATS

Accuracy (Précision) : L'Accuracy mesure le pourcentage de prédictions correctes parmi le total des prédictions. C'est une métrique simple et facile à comprendre pour la classification, mais elle peut être trompeuse en cas de déséquilibre de classes.

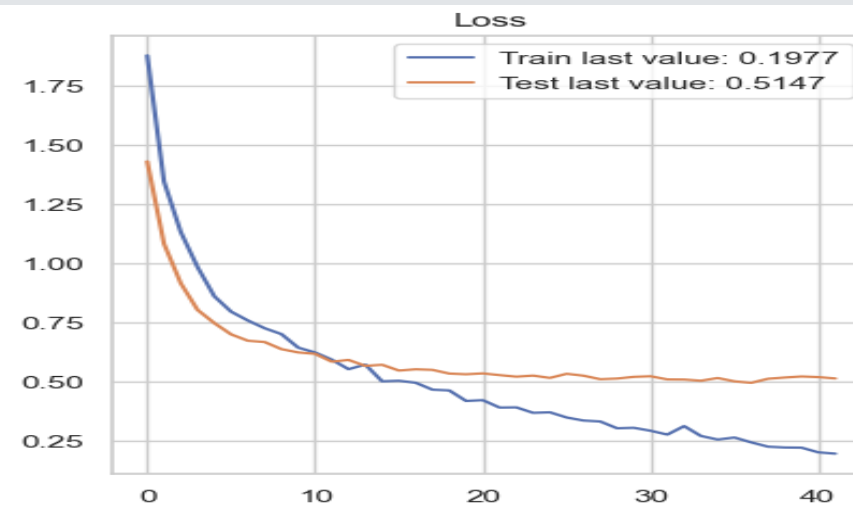
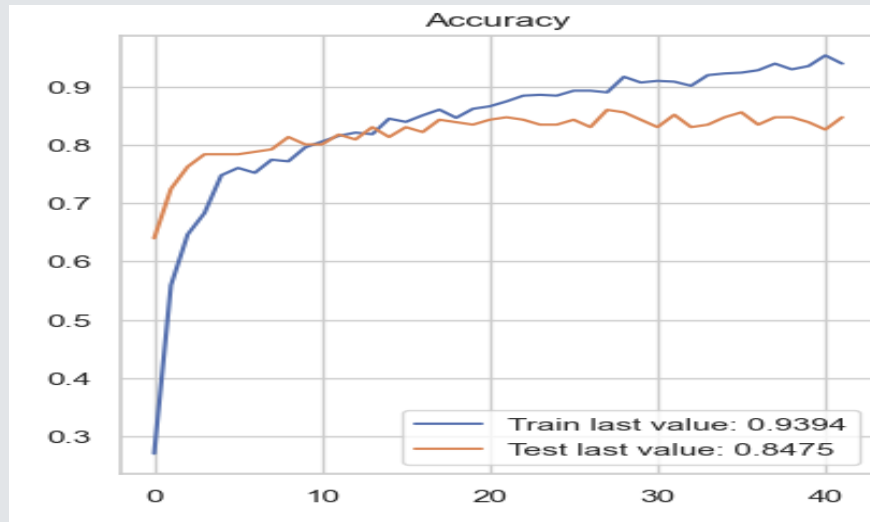
Formule : $Accuracy = \text{Nombre de prédictions correctes} / \text{Nombre total de prédictions}$

Utilisation : Dans ce projet, nous avons calculé l'Accuracy sur les ensembles d'entraînement (Training Accuracy) et de test (Test Accuracy) pour comparer les performances des modèles VGG16 et EfficientNet-B7.

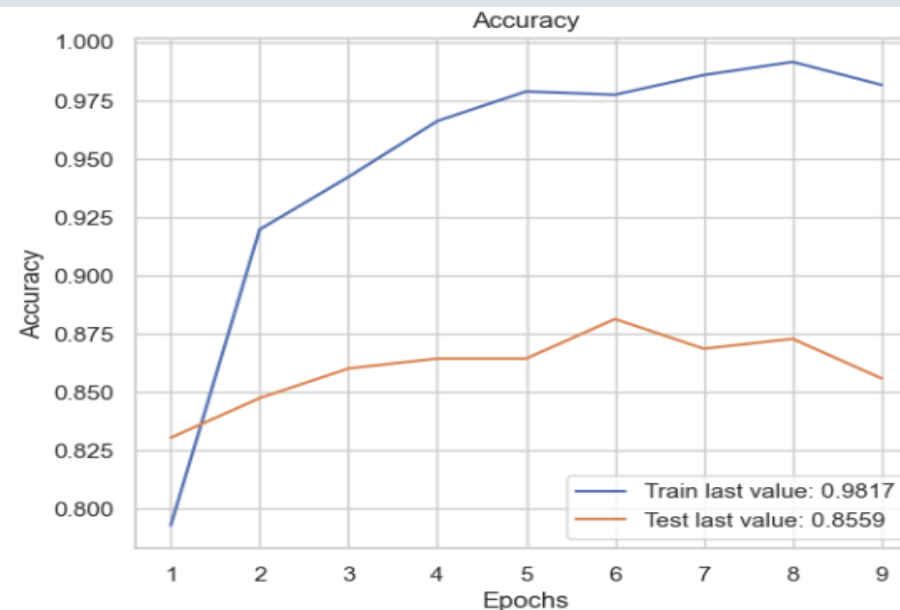
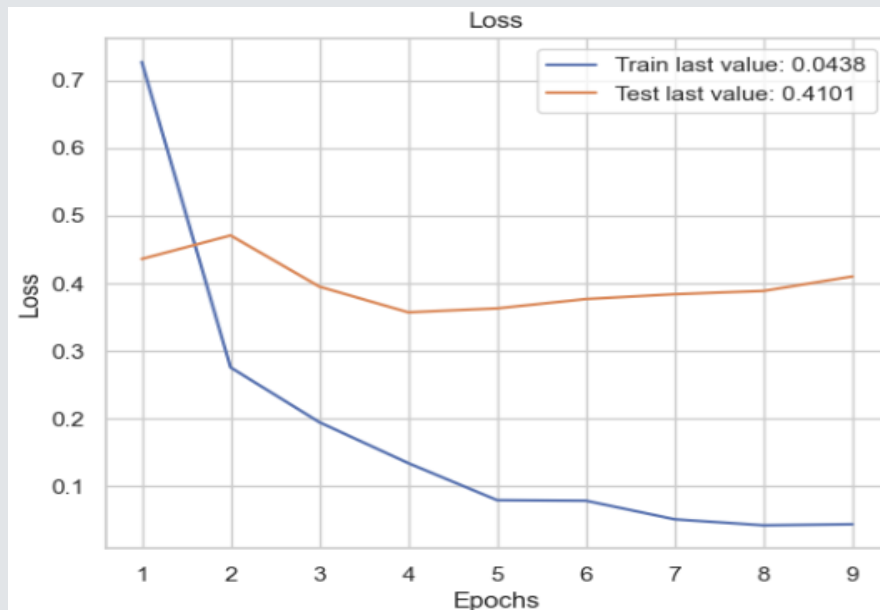
Modèle	Accuracy Entraînement	Accuracy Validation	Accuracy Test
VGG16	93,94 %	83,47 %	84,75 %
EfficientNet-B7	98,17%	86,44 %	12

➤ *EfficientNet-B7 surpasse VGG16 en offrant une meilleure capacité d'apprentissage et de généralisation, avec une accuracy plus élevée tant sur les ensembles de validation que de test, ce qui en fait le modèle le plus performant pour la classification d'images dans notre projet.*

VISUALISATION DES PERFORMANCES DES MODELES

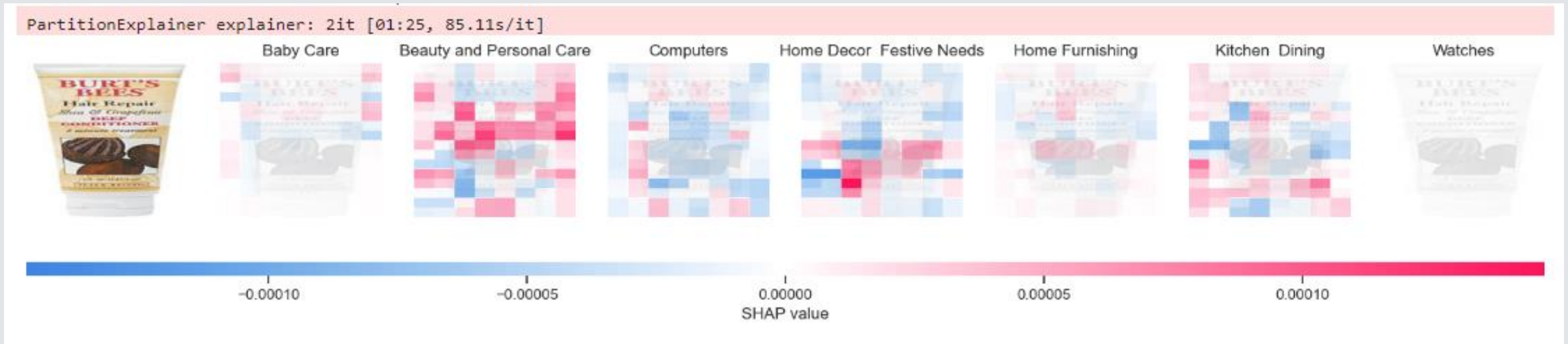


→ VGG16



→ EfficientNet-B7

FAETURES IMPORTANCE : EfficientNet



Heatmap des Valeurs SHAP : Montre les contributions positives (rouge) et négatives (bleu) des zones de l'image pour chaque catégorie.

Méthode SHAP : Permet de visualiser les zones les plus influentes dans les décisions du modèle.

- Les catégories **Kitchen Dining** et **Watches** apparaissent comme les plus influentes dans la prédiction, avec des zones fortement marquées en rouge dans la heatmap, indiquant des contributions positives significatives dans la décision du modèle EfficientNet-B7.

CONCLUSION

❖ **EfficientNet-B7** se positionne comme une solution performante et efficace pour la classification d'images, grâce à sa capacité à équilibrer la profondeur, la largeur et la résolution des réseaux de neurones.

❖ **Le choix entre EfficientNet-B7 et d'autres modèles**, comme VGG16, dépendra des besoins spécifiques du projet, notamment en termes de précision, de rapidité d'entraînement, et d'utilisation des ressources.

❖ **Améliorations potentielles** : accroître la diversité du dataset, ajuster les hyperparamètres pour une meilleure généralisation, et intégrer des méthodes d'augmentation de données pour des résultats encore plus robustes.





Merci

