

Projet #5

Développement d'un Système Complet de Recommandation et de Génération d'Images

Machine learning : Développement de modèles

Directeur de projet en intelligence artificielle - Année 2 - L'École Multimédia

Sommaire

[Contexte](#)

[Présentation](#)

[Contraintes](#)

[Libertés](#)

[Objectifs](#)

[Étapes du projet](#)

[Rendu final](#)

[Evaluations](#)

[Compétences visées](#)

[Conseils](#)

Contexte

Présentation

Vous êtes recruté en tant que spécialiste en Machine Learning pour une entreprise de commerce en ligne qui souhaite améliorer son expérience utilisateur en développant deux fonctionnalités innovantes : un système de recommandation personnalisé et un générateur d'images de produits à partir de descriptions textuelles.

Votre mission consiste à développer, optimiser, et évaluer plusieurs modèles de Machine Learning pour ces tâches, en intégrant des techniques avancées comme les GANs pour la génération d'images et les LLM pour le traitement du langage naturel.

Contraintes

- **Travail en autonomie** : Vous travaillerez seul sur ce projet, ce qui implique une gestion efficace du temps et des ressources.
- **Documentation complète** : Chaque étape du projet doit être clairement documentée
- **Respect des délais** : Le projet doit être livré sous la forme d'une **archive ZIP** et dans les délais imposés sous peine de pénalisation.

Objectifs

Vous devrez avoir réalisé les éléments suivant :

1. **Développer et optimiser un système de recommandation personnalisé** en utilisant des techniques avancées de Machine Learning.
2. **Créer et entraîner un modèle GAN** pour générer des images réalistes de produits à partir de descriptions textuelles.
3. **Appliquer des techniques de réduction de dimensionnalité** et de clustering pour segmenter les utilisateurs et personnaliser les recommandations.
4. **Évaluer et comparer les performances des modèles** pour choisir les plus efficaces.
5. **Documenter et présenter le projet** avec des visualisations et des analyses détaillées des résultats.

Étapes du projet

1. Collecte et Préparation des Données

- **Collecte des données** : Utilisez des datasets disponibles (par exemple, le dataset "Amazon Product Reviews" pour le système de recommandation et un dataset d'images de produits avec descriptions pour les GANs).
- **Préparation des données** : Nettoyez les données, traitez les valeurs manquantes, normalisez les caractéristiques, et préparez les ensembles de données pour l'entraînement des modèles.

2. Développement du Système de Recommandation

- **Implémentation de LSA et Nearest Neighbor** :
 - i. Utilisez l'analyse sémantique latente (LSA) pour comprendre les relations entre les produits et les utilisateurs.
 - ii. Implémentez un modèle k-Nearest Neighbors pour fournir des recommandations personnalisées basées sur les interactions passées des utilisateurs.
- **Optimisation du modèle** :
 - i. Appliquez des techniques de fine-tuning et d'optimisation pour améliorer les performances du système de recommandation.
 - ii. Utilisez des techniques de réduction de dimensionnalité (PCA, LDA) pour améliorer l'efficacité du modèle.

3. Création et Entraînement des GANs

- **Développement d'un GAN pour la génération d'images** :
 - i. Implémentez un réseau antagoniste génératif (GAN) en utilisant TensorFlow/Keras pour générer des images de produits à partir de descriptions textuelles.
 - ii. Entraînez le GAN en optimisant l'architecture et en ajustant les hyperparamètres pour améliorer la qualité des images générées.
- **Applications innovantes des GANs** :
 - i. Explorez l'utilisation des GANs pour des tâches spécifiques, comme la génération d'images de haute qualité ou la super-résolution d'image.
 - ii. Intégrez les GANs dans un pipeline complet qui permet de générer des images directement à partir de descriptions textuelles.

4. Évaluation et Comparaison des Modèles

- **Évaluation des systèmes de recommandation** :
 - i. Mesurez les performances du système de recommandation en utilisant des métriques comme le precision, recall, F1-score, et ROC-AUC.
 - ii. Comparez les performances des différentes approches (LSA, Nearest Neighbor) et choisissez la plus efficace.
- **Évaluation des GANs** :
 - i. Évaluez la qualité des images générées par le GAN à l'aide de métriques comme l'Inception Score et le Frechet Inception Distance (FID).
 - ii. Analysez la capacité des GANs à générer des images réalistes et à répondre aux descriptions textuelles.

5. Documentation et Présentation des Résultats

- **Documentation du projet :**
 - i. Documentez chaque étape du projet, incluant la collecte et la préparation des données, l'implémentation des modèles, l'évaluation des performances, et les conclusions.
- **Visualisation des résultats :**
 - i. Créez des visualisations des performances du système de recommandation et des GANs, incluant des graphes, des matrices de confusion, et des exemples d'images générées.
- **Présentation finale :**
 - i. Préparez une présentation pour expliquer les résultats et les recommandations, incluant des démos des fonctionnalités développées (système de recommandation et génération d'images).

Données

A COMPLÉTER :

- dataset "Amazon Product Reviews" pour le système de recommandation
- dataset d'images de produits avec descriptions pour les GANs

Rendu final

Votre rendu final prendra la forme d'une archive Zip et devra comporter les éléments suivants :

1. **Dataset préparé** : Fichier(s) CSV nettoyé(s) et prêt(s) à être utilisé(s) pour le Machine Learning.
2. **Modèles de Machine Learning** : Scripts Python pour le système de recommandation (LSA, Nearest Neighbor) et le GAN, accompagnés de commentaires et d'explications.
3. **Évaluations des performances** : Fichiers contenant les résultats des évaluations des modèles, incluant les métriques de performance et les visualisations des résultats.
4. **Images générées** : Un ensemble d'images de produits générées par le GAN à partir de descriptions textuelles.
5. **Rapport final** : Un document détaillant toutes les étapes du projet, les résultats des modèles, et les recommandations pour l'entreprise.
6. **Dépôt GitHub** : Dépôt contenant tout le code source, les scripts, la documentation, et l'historique des versions.

Cette archive aura comme titre votre nom et prénom avec votre classe.

Exemple: `arthur_mensch_projet5_AIA02.zip`

Attention : Un rendu non livré ou en retard vous pénalise pour la certification

Evaluations

- **Qualité de la préparation des données** : Pertinence du nettoyage, de la normalisation, et de la préparation des données.
- **Performance du système de recommandation** : Précision des recommandations et efficacité des techniques de modélisation utilisées.
- **Qualité des images générées par le GAN** : Réalisme des images et capacité du modèle à répondre aux descriptions textuelles.
- **Capacité à évaluer et optimiser les modèles** : Utilisation appropriée des métriques et des techniques d'optimisation pour améliorer les performances des modèles.
- **Documentation et gestion du projet** : Exhaustivité et clarté de la documentation, ainsi que l'utilisation appropriée de GitHub pour la gestion du projet.
- **Présentation professionnelle des résultats** : Qualité de la présentation finale, incluant l'explication des choix méthodologiques, l'interprétation des résultats, et les recommandations.

Ce projet d'évaluation permet de valider les compétences des étudiants dans le développement de systèmes de recommandation et de génération d'images, en intégrant des techniques avancées de Machine Learning et en les appliquant à des problématiques réelles. Les étudiants seront évalués sur leur capacité à implémenter, optimiser, et évaluer des modèles complexes, tout en documentant et présentant leurs travaux de manière professionnelle.

Compétences visées

	Modélisation des données
C2.3	Élaborer des modèles de données logiques et physiques (entité-relation, les modèles de données en étoile...) qui correspondent au cahier des charges établi.
C2.4	Concevoir des structures de bases de données adaptées à divers types de données, en tenant compte des performances, de la sécurité, de l'évolutivité, et du volume des données, pour une gestion optimale du Big Data.
C2.5	Déployer des serveurs virtuels dans le cloud ou On-Premise pour l'entraînement des algorithmes d'Intelligence Artificielle, en vue de gérer efficacement un large volume de données.
	Optimisation des performances de l'infrastructure
C2.6	Augmenter la puissance de calcul à travers la conception de clusters de serveurs, afin de permettre l'entraînement d'algorithmes d'Intelligence Artificielle, le stockage de données volumineuses ou encore l'accueil de trafic massif sur une application.
	Développement des pipelines de données
C3.1	Concevoir un système de gestion de données temps réel adapté aux contraintes et normes opérationnelles de l'entreprise, pour gérer efficacement la vitesse, le volume des flux, et la typologie des données.
C3.2	Établir un pipeline de données à travers des processus ETL/ELT pour le transfert et la transformation des données entre différentes bases, en utilisant des outils de programmation, afin de répondre aux spécifications du cahier des charges.
	Construction de la solutions d'IA
C4.1	Rédiger un cahier des charges pour la solution d'Intelligence Artificielle, afin de répondre aux besoins techniques et économiques de l'organisation, en tenant compte de l'accessibilité pour les personnes en situation de handicap.
C4.2	Créer un algorithme d'Intelligence Artificielle adapté aux données d'entraînement et conforme aux spécifications du cahier des charges, en veillant à répondre aux besoins spécifiques, notamment en termes d'accessibilité.

Conseils

- Bien prendre le temps d'analyser le brief et comprendre le client
- Organisez-vous et planifiez votre travail : donnez vous des objectifs intermédiaires
- Planifiez des sessions de travail régulière
- Ne jamais être trop ambitieux
- Faites directement les documents du livrable
- Mettez en oeuvre les bonnes pratiques vues en cours