

Projet : Optimisation de l'allocation de la bande passante par apprentissage automatique

Description du problème:

Les fournisseurs de services de télécommunications sont constamment confrontés à la nécessité d'optimiser la bande passante pour répondre efficacement à la demande fluctuante des utilisateurs. La gestion inefficace de la bande passante peut entraîner une dégradation de la qualité de service, des retards ou même des interruptions de service, ce qui affecte la satisfaction client et peut entraîner des pertes financières significatives.

Contexte spécifique:

Avec l'augmentation exponentielle des données mobiles et de l'utilisation d'internet reportée par les statistiques de l'OCDE, ainsi que le développement de nouvelles applications gourmandes en données comme le streaming vidéo, les jeux en ligne, et la réalité augmentée, il devient crucial de développer des solutions plus dynamiques et intelligentes pour la gestion de la bande passante.

Objectif Général

Ce projet vise à permettre aux étudiants de mettre en pratique les concepts enseignés en cours en appliquant des méthodes d'analyse et de modélisation des données. Les étudiants travailleront en groupe de 3 et utiliseront les jeux de données définis dans le sites :

<https://stats.oecd.org/Index.aspx?DataSetCode=TEL>

L'objectif est d'explorer, nettoyer, analyser et développer un modèle de machine learning capable de prédire la demande de bande passante en temps réel et d'ajuster automatiquement la répartition des ressources réseau pour maximiser l'efficacité et la qualité de service en suivant les étapes décrites ci-dessous.

Tâches à Réaliser

1. Sélection et Préparation des Données

Chaque groupe devra :

1. Choisir au moins deux caractéristiques du jeu de données fournies.
Effectuer un nettoyage des données :

- a. Identifier et traiter les valeurs manquantes.
 - b. Gérer les anomalies (valeurs extrêmes ou incohérentes).
 - c. Supprimer ou fusionner les doublons si nécessaire.
2. Réaliser une analyse exploratoire :
 - a. Décrire les données sélectionnées à l'aide de statistiques descriptives.
 - b. Créer des visualisations pertinentes pour comprendre les relations entre les variables.
 - c. Normaliser ou transformer les données si nécessaire (par exemple : centrage, échelle logarithmique).

2. Application des Méthodes Apprises en Cours

Les groupes appliqueront les algorithmes suivants et interpréteront leurs résultats :

- K-Means : Réaliser un regroupement des données en clusters.
- Analyse hiérarchique des clusters (HCA) : Identifier les hiérarchies dans les données.
- K-Nearest Neighbors (KNN) : Créer un modèle de classification (une cible sera choisie par le groupe).
- Régression linéaire simple et multiple : Prédire une variable numérique à partir des caractéristiques sélectionnées.
- Régression logistique : Modéliser des relations entre des variables et une cible binaire ou catégorielle.

Pour chaque méthode, les étudiants doivent :

- Expliquer pourquoi cette méthode est utilisée.
- Documenter les paramètres choisis.
- Interpréter les résultats obtenus.

3. Recherche et Application des Séries Temporelles

Chaque groupe doit :

- Effectuer une recherche sur les méthodes d'analyse des séries temporelles:
 - ARIMA (Auto-Regressive Integrated Moving Average).
 - LSTM (Long Short-Term Memory).
- Choisir une méthode d'analyse temporelle parmi celles recherchées.
- Appliquer cette méthode à un sous-ensemble des données (par exemple, évaluer les tendances ou cycles dans une caractéristique temporelle).
- Interpréter les résultats de l'analyse temporelle.

4. Production des Livrables

Les groupes devront soumettre la veille de leur dernier TD:

- Un fichier.py contenant le code ainsi qu'un rapport(pdf):
 - L'introduction présentant les caractéristiques sélectionnées et leur pertinence.
 - Les étapes de nettoyage et préparation des données.
 - Une présentation des méthodologies appliquées pour chaque algorithme.
 - Les résultats, les visualisations et interprétations associées.
 - La recherche sur les séries temporelles et leur application.
 - Une conclusion sur les leçons tirées et les améliorations potentielles.
- Une présentation orale (soutenance de 10 minutes pendant le dernier TD) comprenant :
 - Une synthèse des étapes réalisées.
 - Les résultats clés.
 - Les principaux défis rencontrés et solutions apportées.

Critères d'Évaluation

Critères	Poids
Qualité de la préparation des données	20%
Pertinence et mise en œuvre des méthodes	30%
Recherche et application des séries temporelles	20%
Clarté et qualité du rapport écrit	20%
Présentation orale	10%

Déroulement

- Les groupes travailleront sur leur projet durant les deux dernières séances avant la séance finale, consacrée aux présentations.
- Un suivi sera organisé pour vérifier l'avancement des groupes pendant les séances de TD.
- La date de soumission des livrables est prévue avant la veille du jour de la présentation.

Bonne chance à tous !