



Sujet 13

Machine learning temps réel sur flux de données

Réalisé par :

- Khadija CHAHIDI
- Siham HAFSI
- Nassim EI BOUGRINI
- Ilham Nachour
- Achraf AGOUMI
- HAYAT RAHMOUN

Encadré par :

- M. Thierry Bertin

Plan

01

Définition

02

Défis

03

Approches

Applications

04

Outils

05

01

Définition



Machine Learning temps réel sur flux de données:



02

Défis



Défis :



Vitesse de
traitement



Evolution des
concepts



Gestion de volume
des données



Latence et temps
de réponse



Qualité des
données

03

Approches du Machine Learning en
temps réel sur flux de données

Approches du Machine Learning
en temps réel sur flux de
données



Apprentissage incrémental

L'apprentissage incrémental permet de mettre à jour le modèle progressivement à mesure que de nouvelles données arrivent.

Approches du Machine Learning
en temps réel sur flux de
données



Fenêtres glissantes

Les fenêtres glissantes sont utilisées pour diviser le flux de données en petits sous-ensembles (fenêtres) qui se chevauchent généralement.

Approches du Machine Learning
en temps réel sur flux de
données



Mini-batch learning

Le mini-batch learning
consiste à regrouper les
données en petits lots de
taille fixe pour les mises à
jour du modèle.

Approches du Machine Learning
en temps réel sur flux de
données



Algorithmes adaptatifs

Les algorithmes adaptatifs ajustent automatiquement leurs paramètres en fonction des variations des données.

Approches du Machine Learning
en temps réel sur flux de
données



Techniques de prétraitement en continu

Les techniques de
prétraitement en continu,
comme PCA incrémental,
sont utilisées pour réduire la
complexité des données en
temps réel.

Approches du Machine Learning
en temps réel sur flux de
données



Méthodes d'ensemble

Les méthodes d'ensemble
combinent les prédictions de
plusieurs modèles pour
améliorer la précision et la
robustesse.




04

Applications



Prévision météorologique à l'aide de l'apprentissage automatique


Les prévisions météorologiques en temps réel sont une application de machine learning qui utilise des modèles informatiques basés sur des données historiques et en temps réel pour prédire les conditions météorologiques futures. Les algorithmes de machine learning analysent de vastes ensembles de données provenant de capteurs météorologiques, de radars et de satellites pour identifier les schémas et les tendances météorologiques. Ces prévisions sont essentielles pour la planification des activités extérieures, la sécurité publique, l'agriculture, l'aviation, la navigation maritime, la gestion des ressources en eau, entre autres. Cependant, il convient de noter que les prévisions météorologiques restent sujettes à l'incertitude en raison de la complexité du système atmosphérique et des variations naturelles du climat. Les prévisions en temps réel fournissent des indications importantes, mais elles ne sont pas infaillibles et doivent être utilisées avec prudence.





Reconnaissance d'objets en temps réel dans les vidéos ou les images

La reconnaissance d'objets en temps réel dans les vidéos ou les images est une application de machine learning qui utilise des algorithmes pour détecter, identifier et suivre automatiquement des objets spécifiques dans des flux vidéo ou des images. Cette technologie a des applications variées, telles que la surveillance de la sécurité, les voitures autonomes, la réalité augmentée, les systèmes de caisse automatique, la médecine, et bien d'autres encore. Elle contribue à rendre les systèmes informatiques plus intelligents et interactifs avec le monde réel.



Applications du Machine
Learning en temps réel sur flux
de données



Détection d'anomalies en temps réel

La détection d'anomalies en temps réel avec le machine learning est une technique d'analyse de données qui vise à identifier des schémas anormaux ou des événements inhabituels dans des flux de données continus au fur et à mesure qu'ils se produisent.

Applications du Machine
Learning en temps réel sur flux
de données

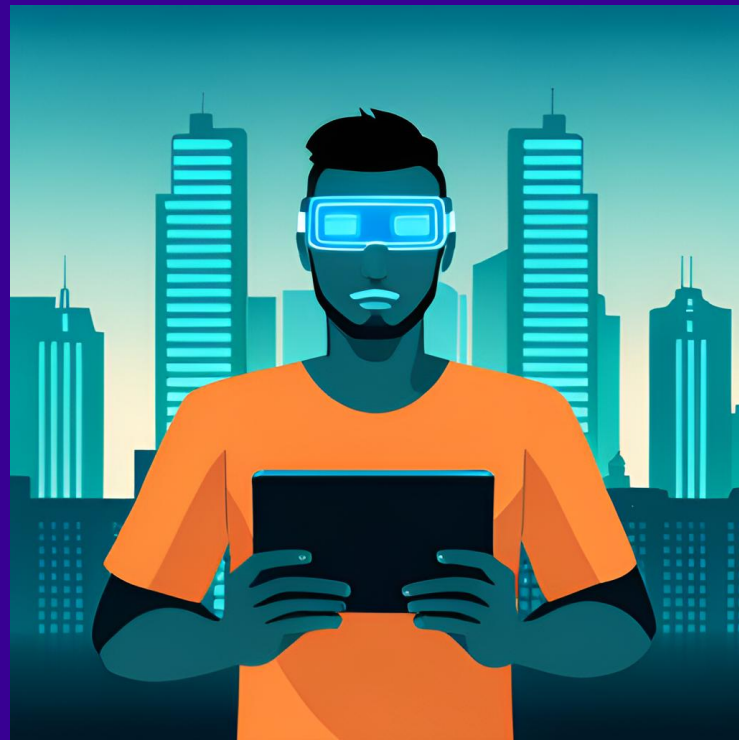


Détection d'URL malveillantes en temps réel

La détection d'URL malveillantes en temps réel à l'aide de l'apprentissage automatique vise à identifier les URL ou liens web qui redirigent vers des sites malveillants ou dangereux. Cette approche permet de protéger les utilisateurs en temps réel contre les attaques.

05

Outils



Langages de programmation

Python

Langage de programmation interprété et polyvalent, largement utilisé dans le domaine du machine learning.

Java

Langage de programmation orienté objet, réputé pour sa fiabilité et sa capacité à gérer des systèmes d'entreprise, y compris ceux impliquant du traitement en temps réel.

Scala

Langage de programmation s'exécutant sur la machine virtuelle Java (JVM), combinant les fonctionnalités de la programmation orientée objet et fonctionnelle.

Bases de données

MongoDB

Base de données NoSQL orientée document, adaptée pour le stockage et la gestion de données.

Apache Cassandra

Base de données NoSQL distribuée, conçue pour la haute disponibilité et la scalabilité linéaire, idéale pour gérer de gros volumes de données en temps réel.

Apache HBase

Base de données NoSQL distribuée, utilisant un modèle de données colonne, offrant une gestion rapide et évolutive des données .

TimescaleDB

Base de données NoSQL distribuée, utilisant un modèle de données colonne, offrant une gestion rapide et évolutive des données chronologiques.

Bibliothèques de machine learning

TensorFlow

open-source pour l'apprentissage profond et la résolution de problèmes de classification, régression, NLP et vision par ordinateur.

Scikit-learn

offrant des algorithmes supervisés et non supervisés ainsi que des outils pour la préparation des données et l'évaluation des modèles.

Keras

Interface de programmation en Python permettant de créer facilement des modèles d'apprentissage profond, souvent utilisée en conjonction avec des bibliothèques de calcul numérique comme TensorFlow ou Theano.

Apache Spark MLlib

intégrée dans le framework de traitement distribué Apache Spark, proposant divers algorithmes de machine learning pour le traitement de données en temps réel et par lots.

Systemes de traitement en temps réel

Apache Kafka

Plateforme de streaming distribuée pour la gestion de flux de données en temps réel, utilisée comme file d'attente de messages et système de journalisation.

Apache Flink

Moteur de traitement de flux de données en temps réel et par lots, avec faible latence et tolérance aux pannes.

Apache Spark Streaming

Moteur de traitement de flux de données en temps réel et par lots, avec faible latence et tolérance aux pannes.

Apache Storm

Système de traitement de flux de données en temps réel, conçu pour la scalabilité et utilisé pour les applications nécessitant une faible latence et un traitement continu d'événements.

Outils de visualisation

Tableau

Plateforme de visualisation de données permettant de créer des graphiques interactifs et des tableaux de bord visuels.

Grafana

Outil open-source de visualisation et de monitoring, spécialisé dans les métriques et les données temps réel.

Kibana

Interface utilisateur pour la plateforme ELK (Elasticsearch, Logstash, Kibana), utilisée pour la visualisation et l'analyse de données de type log en temps réel.

Power BI

Plateforme de Business Intelligence développée par Microsoft, permettant de créer des rapports interactifs et des tableaux de bord visuels.



Merci!

Avez-vous des questions?