

Résumé et points forts

Statut : Traduit automatiquement de Anglais

Traduit automatiquement de Anglais

Nous vous félicitons ! Vous avez terminé ce module. Dans ce module, vous avez appris que :

- Apache Spark Structured Streaming prend en charge le traitement des données en continu à l'aide de Spark SQL.
- L'extraction et la transformation des caractéristiques sont des composants essentiels du prétraitement des données.
- L'extraction de caractéristiques implique la sélection ou la dérivation de caractéristiques significatives à partir de données brutes à des fins d'analyse ou de modélisation.
- MLib fournit diverses techniques d'extraction de caractéristiques telles que la tokenisation, TF-IDF, Word2Vec, et PCA.
- La transformation des caractéristiques consiste à convertir les caractéristiques extraites dans un format approprié pour une analyse ultérieure.
- Spark propose des techniques de transformation de caractéristiques couramment utilisées telles que Scaling and normalization, One-hot encoding, StringIndexer, et PCA.
- Les pipelines d'apprentissage automatique automatisent et structurent le processus d'apprentissage automatique de bout en bout.
- Ils impliquent des étapes interconnectées telles que l'ingestion de données, le nettoyage, le prétraitement, etc.
- Les avantages des modèles d'apprentissage automatique incluent la reproductibilité, l'évolutivité, l'automatisation, la flexibilité, et bien d'autres encore.
- Spark fournit des outils puissants pour construire et déployer des modèles ML à grande échelle.
- Ses composants clés simplifient les tâches de ML, offrant des avantages tels que l'évolutivité, la performance et la facilité d'utilisation.
- La persistance du modèle est un concept important de l'apprentissage automatique qui implique la sauvegarde d'un modèle entraîné sur le disque en vue d'une utilisation ultérieure.
- La persistance des modèles est un composant essentiel des pipelines d'apprentissage automatique.
- En permettant la sauvegarde et la réutilisation des modèles formés, elle améliore l'efficacité, l'évolutivité et la flexibilité des systèmes d'apprentissage automatique.

- Ses avantages contribuent collectivement à améliorer la productivité, la reproductibilité et la collaboration dans le domaine de l'apprentissage automatique.