

Checklist de la maintenance des données

Bienvenue dans l'ère de l'intelligence artificielle (IA), où votre façon de travailler repose sur des technologies consommatrices de données telles que le machine learning et le deep learning. Pour profiter de ces nouveaux outils d'IA, vous devez vous assurer que la « maison » des données de votre organisation est en ordre.

Voici une liste de contrôle pour vous préparer à la transition vers une « maison de données » propre, divisée en deux phases clés de maintenance : la formation et l'inférence

Suivez ces étapes pour vous aider à maîtriser l'IA. Pour plus de renseignements sur la façon de faire passer l'IA de la phase de validation de concept à la phase de mise en production et évolution, consultez le rapport IDC, [Accélérer et rendre opérationnels et les déploiements IA à l'aide d'une infrastructure optimisée pour l'IA.](#)

Formation

Dans la phase de formation de la préparation à l'IA, vous développerez des algorithmes pour comprendre un ensemble de données. Votre principale préoccupation sera de rassembler les données existantes et d'utiliser l'IA pour apprendre une nouvelle capacité.

- Déterminez le problème métier spécifique que vous souhaitez résoudre à l'aide de l'IA (commencez par des petits projets pour vous aider à apprendre)
- Localisez les données qui peuvent vous aider à résoudre ce problème auprès de sources pertinentes (elles ne seront probablement pas toutes placées au même endroit)
- Préparez vos données avec des balises de métadonnées pour réduire considérablement le temps requis pour trouver des données pertinentes
- Assurez-vous que vos données sont correctement synchronisées et liées sur tous les ensembles que vous utiliserez (synchronisées dans le temps notamment)
- Identifiez toutes les données client sensibles et autres données privées pour vous assurer de garantir absolument leur sécurité et de respecter toutes les exigences réglementaires et de gouvernance appropriées (le processus de balisage des métadonnées peut vous y aider)
- Choisissez le bon environnement de développement pour le type de données que vous utilisez et la façon dont elles seront formatées (par ex. les données image, vidéo, texte non structuré et audio ont généralement leur propre type d'environnement)
- Extrayez des ensembles de données de votre référentiel et apportez-les dans votre environnement de développement
- Divisez vos données en deux groupes pour vous aider à améliorer votre processus de développement de modèle (placez un ensemble dans un dossier appelé « formation » et un autre ensemble dans un dossier appelé « test »)
- Maintenez la traçabilité des données en effectuant le suivi de l'endroit et des sources d'où proviennent vos données (envisagez d'utiliser des outils qui peuvent vous aider à automatiser le processus)
- Effectuez des tâches d'hygiène élémentaire pour préparer les données à la création d'un modèle (par exemple, complétez les informations manquantes ou prévoyez la suppression des entrées nulles)
- Utilisez un exemple de sous-ensemble de données pour lequel vous connaissez déjà la réponse à l'activité de prédiction (cela s'appelle un « ensemble de formation ») et identifiez toutes les étapes de traitement préalable nécessaires pour préparer les données à la réalisation d'une prédiction
- Utilisez votre connaissance de cet ensemble de formation pour calculer les scores de précision qui peuvent vous permettre d'appliquer en toute confiance le même modèle aux nouvelles données pour lesquelles le modèle n'a jamais été explicitement formé

Inférence

Une fois que vous aurez développé un modèle qui permet de résoudre votre problème métier, vous passerez de la formation à l'inférence. Dans cette phase, vous prenez ce modèle couronné de succès et l'appliquez aux nouvelles données, ce qui exige également une certaine maintenance des données.

- Placez votre modèle d'IA près de vos données pour réduire la latence, réduire les exigences de bande passante et améliorer les performances globales du modèle
- Développez un processus de pipeline de données efficace et appliquez l'étiquetage des métadonnées à vos données au fur et à mesure de leur entrée, afin que vos nouvelles données puissent être rassemblées et utilisées pour améliorer le modèle dans les étapes suivantes
- Insérez des balises dans les données de façon liée et synchronisée (par exemple, si les données sont classées par ordre chronologique, vous pouvez synchroniser les ensembles de données ou les lier en choisissant un champ, par exemple le nom d'un client, sur toutes les données qui entrent)
- Développez un plan de stockage pour les données à long terme, afin de définir la façon dont vous en gérerez le volume et la vitesse, à mesure qu'elle entrent et que vous les archivez
- Envisagez de recruter un Chief Data Officer pour effectuer la maintenance de la gestion des données pour les futurs projets d'IA, de deep learning et autres projets axés sur les données