

Mis à jour le 09/01/2025

S'inscrire

Formation BentoML

2 jours (14 heures)

Présentation

Notre formation BentoML vous enseignera la maîtrise et la compréhension de la création, la gestion et le déploiement de modèles Machine Learning. Approfondissez vos connaissances et vos compétences dans la production, tout en simplifiant les processus.

Avec notre formation, entrer dans l'ère du "déploiement intelligent" BentoML est la solution idéale pour simplifier le déploiement de vos modèles de machine learning. Cet outil open source vous permet de maîtriser vos modèles en API tout en garantissant une gestion sûre et flexible de vos projets.

Durant cette formation, vous découvrirez comment moderniser les déploiements de modèles ML en transformant vos systèmes applicatifs pour permettre une mise à l'échelle horizontale efficace et une intégration fluide dans vos infrastructures existantes.

Comme pour toutes nos formations, celle-ci vous présentera la toute dernière version de [BentoML V1.3.19](#)

Objectifs

- Comprendre l'intérêt et les caractéristiques de BentoML pour le déploiement de modèles ML
- Savoir transformer des modèles ML en services applicatifs prêts pour la production
- Maîtriser les outils de packaging et de gestion des modèles avec BentoML
- Permettre le passage à l'échelle automatique des services de prédiction
- Optimiser la gestion des dépendances et des ressources pour des modèles lourds
- Développer un monitoring avancé des performances des modèles en production
- Automatiser le déploiement et les mises à jour des versions des modèles

Public visé

- DevOps
- Développeurs
- Data Scientists
- Machine Learning Engineers

Pré-requis

- Bonne compétence en Python
- Connaissances en Machine learning

Programme de formation BentoML

Introduction à BentoML

- Qu'est-ce que BentoML ?
- Analyse de l'architecture BentoML
- Pré-requis techniques
- Installation de BentoML et configuration de l'environnement
- Interface CLI et bibliothèques associées

Le Packaging des Modèles

- Packaging des modèles scikit-learn, TensorFlow, et PyTorch avec BentoML
- Gestion des dépendances dans un environnement Python avec des bibliothèques spécifiques
- Introduction à BentoBundle

Création et déploiement d'un service simple

- Création d'un service REST/GRPC autour d'un modèle ML
- Ajout de pipelines de prétraitement et post-traitement
- Test local des services
- Debugging des erreurs
- Déploiement local via Docker
- Présentation des intégrations avec Kubernetes

Déploiement sur le cloud

- Déploiement sur AWS, GCP, et Azure
- BentoML avec Kubernetes pour le scaling
- Configuration de Docker pour du déploiement multi-environnement

- Mise en place de métriques avec Prometheus et Grafana
- Analyse des performances
 - latence
 - temps de réponse
- Gestion des versions des modèles en production

Optimisation des performances et cas pratiques

- Scaling horizontal et vertical
- Optimisation des API pour les charges lourdes
 - Batching
 - Multi-threading
- Utilisation des GPU pour les modèles lourds
- Intégration d'un modèle de recommandation dans une application web
- Mise en production d'un modèle de classification avec un pipeline complet
- Gestion des erreurs courantes
 - Dépendances
 - Incompatibilités

Sociétés concernées

Cette formation s'adresse à la fois aux particuliers ainsi qu'aux entreprises, petites ou grandes, souhaitant former ses équipes à une nouvelle technologie informatique avancée ou bien à acquérir des connaissances métiers spécifiques ou des méthodes modernes.

Positionnement à l'entrée en formation

Le positionnement à l'entrée en formation respecte les critères qualité Qualiopi. Dès son inscription définitive, l'apprenant reçoit un questionnaire d'auto-évaluation nous permettant d'apprécier son niveau estimé sur différents types de technologies, ses attentes et objectifs personnels quant à la formation à venir, dans les limites imposées par le format sélectionné. Ce questionnaire nous permet également d'anticiper certaines difficultés de connexion ou de sécurité interne en entreprise (intraentreprise ou classe virtuelle) qui pourraient être problématiques pour le suivi et le bon déroulement de la session de formation.

Méthodes pédagogiques

Stage Pratique : 60% Pratique, 40% Théorie. Support de la formation distribué au format numérique à tous les participants.

Organisation

Le cours alterne les apports théoriques du formateur soutenus par des exemples et des séances de réflexions, et de travail en groupe.

Validation

À la fin de la session, un questionnaire à choix multiples permet de vérifier l'acquisition correcte des compétences.

Sanction

Une attestation sera remise à chaque stagiaire qui aura suivi la totalité de la formation.