

Programme de la formation MLE pour Fiducré

Histoire du Machine Learning et contexte du Big Data

- Remplacer à leur échelle les concepts d'Intelligence Artificielle, apprentissage automatique (machine learning)...
- Le lien avec les mathématiques, les statistiques (inférentielles), le data mining et la data science.
- Passer de l'analyse descriptive à l'analyse prédictive puis prescriptive.
- Les applications du Machine Learning (moteurs de recherche, détection des spams, lecture des chèques).
- La typologie des algorithmes de Dominique CARDON.
- La communauté Data Science et les challenges Kaggle (ex. de Netflix).

Etude de cas

Etudes d'applications concrètes du Machine Learning (moteurs de recherche, détection des spams, lecture des chèques).

Les outils du marché pour le traitement de la donnée et le Machine Learning

- Les logiciels traditionnels (SAS, SPSS, Stata...) et leur ouverture à l'Open Source.
- Choisir entre les deux leaders Open Source : Python et R.
- Plateformes Cloud (Azure, AWS, Google Cloud Platform) et solutions SaaS (IBM Watson, Dataiku).
- Nouveaux postes en entreprises : data engineer, data scientist, data analyst, etc.
- Associer les bonnes compétences à ces différents outils.
- Les API en ligne (IBM Watson, Microsoft Cortana Intelligence...).
- Les chatbots (agents conversationnels).

Démonstration

Démonstration d'un chatbot (agent conversationnel) et d'Azure Machine Learning.

Les différents types d'apprentissage en Machine Learning

- Apprentissage supervisé : répéter un exemple.
- Apprentissage non supervisé : découvrir les données.
- Online (Machine) Learning par opposition aux techniques batch.
- Reinforcement learning : optimisation d'une récompense.
- Autres types d'apprentissage (par transfert, séquentiel, actif...).
- Illustrations (moteurs de recommandation...).

Démonstration Démonstrations sur les différents types d'apprentissage Machine Learning possibles.

Les algorithmes du Machine Learning

- Régression linéaire simple et multiple. Limites des approches linéaires.
- Régression polynomiale (LASSO). Séries temporelles.
- Régression logistique et applications en scoring.
- Classification hiérarchique et non hiérarchique (KMeans).
- Classification par arbres de décision ou approche Naïve Bayes.
- Random Forest (développement des arbres de décision).
- Gradient Boosting. Réseaux de neurones. Machine à support de vecteurs.
- Deep Learning : exemples et raisons du succès actuel.
- Text Mining : analyse des corpus de données textuelles.

Démonstration

Démonstration des différents algorithmes de base sous R ou Python.

Procédure d'entraînement et d'évaluation des algorithmes

- Séparation du jeu de données : entraînement, test et validation.
- Techniques de bootstrap (bagging).
- Exemple de la validation croisée.
- Définition d'une métrique de performance.
- Descente de gradient stochastique (minimisation de la métrique).
- Courbes ROC et de lift pour évaluer et comparer les algorithmes.
- Matrice de confusion : faux positifs et faux négatifs.

Démonstration

Démonstration du choix du meilleur algorithme.

Traitement automatique du texte et de la langue

- Comment traiter la langue ? Une donnée très structurée (Grammaire, Sémantique)
- Le résumé automatique
- Mise en relation de documents et extraction d'informations importantes
- La traduction automatique
- La génération de texte
- Analyse audio : Reconnaître la voix et les locuteurs
- Identifier les grands acteurs du traitement textuel

Démonstration

Démonstration d'outils de résumé automatique et de génération de texte

Mise en production d'un algorithme de Machine Learning

- Description d'une plateforme Big Data.
- Principe de fonctionnement des API.
- Du développement à la mise en production.
- Stratégie de maintenance corrective et évolutive.
- Evaluation du coût de fonctionnement en production.

Démonstration

Démonstration d'API de géolocalisation et d'analyse de sentiments.

Le futur de l'IA : Les grands changements récents et leurs impacts

- Les nouveaux algorithmes génératifs : ChatGPT, Bard, Dall-E, Midjourney. . .
- Cas d'utilisations possibles
- Les milieux professionnels impactés
- Création ou perroquet statistique
- Les limites des modèles génératifs ; quand la machine devient mythomane

Démonstration

Utilisation de ChatGPT pour différents usages d'application

Aspects éthiques et juridiques liés à l'Intelligence Artificielle

- Missions de la CNIL et évolutions à venir.
- Question du droit d'accès aux données personnelles.
- Question de la propriété intellectuelle des algorithmes.
- Nouveaux rôles dans l'entreprise : Chief Data Officer et Data Protection Officer.
- Question de l'impartialité des algorithmes.
- Attention au biais de confirmation.
- Les secteurs et les métiers touchés par l'automatisation.

Réflexion collective

Réflexion en commun pour identifier les clés de réussite.