



Agro Data Scientist : De la Biostatistique au Deep Learning pour le Vivant

Lien :

<https://afrique-formation.com/formation/agro-data-scientist-de-la-biostatistique-au-deep-learning-pour-le-vivant>

 DURÉE
10 jours (70h)

 RÉFÉRENCE
AA05

 CATÉGORIE
**Agriculture Intelligente
et Agriecologie**

🎯 OBJECTIFS DE LA FORMATION

À l'issue de cette formation, vous serez capable de :

- ✓ Maîtriser les langages R et Python pour le traitement de données biologiques massives
- ✓ Concevoir des plans d'expériences robustes et en analyser la significativité statistique
- ✓ Développer des modèles prédictifs (Machine Learning) adaptés à la variabilité du vivant
- ✓ Traiter des données spectrales (NIRS) pour le contrôle qualité non destructif
- ✓ Analyser et cartographier la variabilité spatiale des sols et des rendements (Geostats)
- ✓ Utiliser le Deep Learning pour automatiser le phénotypage par imagerie
- ✓ Créer des tableaux de bord interactifs pour l'aide à la décision agronomique
- ✓ Intégrer des données hétérogènes (Climat, Sol, Plante, Genotype) dans une approche systémique
- ✓ Évaluer la performance et la robustesse des modèles avant leur déploiement terrain
- ✓ Communiquer efficacement les résultats d'analyses complexes à des parties prenantes métier

👥 POUR QUI ?

- ✓ Ingénieurs agronomes souhaitant acquérir une double compétence en Data Science
- ✓ Data Scientists généralistes désirant se spécialiser dans les sciences du vivant
- ✓ Chercheurs et doctorants en agronomie, écologie ou génétique végétale
- ✓ Responsables R&D dans l'industrie semencière ou phytosanitaire
- ✓ Analystes de données au sein d'instituts techniques ou de chambres d'agriculture
- ✓ Responsables Qualité et Laboratoire en industrie agroalimentaire (volet Chimométrie)
- ✓ Consultants en agriculture de précision et agriculture numérique
- ✓ Bio-informaticiens cherchant à élargir leur champ d'action vers l'échelle parcelle



Programme détaillé

1 / Fondamentaux de la gestion de données agronomiques

- Structuration et nettoyage de bases de données complexes (Data Wrangling avec Pandas/Tidyverse)
- Gestion des valeurs aberrantes et données manquantes dans les essais biologiques
- Interopérabilité des formats (CSV, JSON, NetCDF) et bases de données SQL pour l'agronomie

2 / Biostatistiques et inférence classique

- Conception de plans d'expériences (Blocs randomisés, Split-plot, Carrés latins)
- Analyse de variance (ANOVA) et tests post-hoc pour la comparaison de variétés
- Régressions linéaires simples et multiples pour modéliser les interactions GxE (Genotype x Environnement)

3 / Analyse de données multidimensionnelles (Exploratoire)

- Analyse en Composantes Principales (ACP) pour la caractérisation phénotypique
- Analyse Factorielle des Correspondances (AFC) pour les données qualitatives
- Classification hiérarchique (CAH) et K-means pour le zonage ou le typage variétal

4 / Introduction au Machine Learning pour le vivant (Supervisé)

- Algorithmes d'arbres de décision et Random Forests pour la prédiction de rendements
- Support Vector Machines (SVM) pour la classification binaire (Malade/Sain)
- Évaluation des modèles : Matrice de confusion, ROC-AUC et validation croisée (Cross-validation)

5 / Machine Learning pour le vivant (Non-supervisé et Avancé)

- Gradient Boosting (XGBoost, LightGBM) appliqué aux données agro-climatiques
- Sélection de variables (Feature importance) pour identifier les facteurs limitants
- Clustering DBSCAN pour la détection d'anomalies spatiales dans les parcelles

6 / Series temporelles et modélisation climatique

- Analyse de tendances et saisonnalité des données météorologiques
- Modèles ARIMA et LSTM (Long Short-Term Memory) pour la prévision de récolte
- Calcul d'indices bioclimatiques cumulés (GDD, stress thermique)

7 / Chimétrie et analyse spectrale (Agroalimentaire)

- Pré-traitement des spectres proche infrarouge (NIRS) : Derivation, SNV, MSC
- Régression PLS (Partial Least Squares) pour la prédiction de la qualité (Protéines, Brix)
- Discrimination de l'origine géographique des produits par analyse discriminante (PLS-DA)

8 / Géostatistiques et analyse spatiale

- Analyse de la structure spatiale par variographie expérimentale
- Interpolation par Krigage pour la cartographie de la fertilité des sols
- Gestion de l'autocorrélation spatiale dans les modèles agronomiques

9 / Deep Learning et Vision par Ordinateur (Computer Vision)

- Architectures de réseaux de neurones convolutifs (CNN) pour l'analyse d'images foliaires
- Détection d'objets (YOLO) pour le comptage de fruits ou d'adventices
- Segmentation sémantique pour l'estimation de la surface foliaire (LAI)

10 / Analyse de données génomiques et sélection assistée

- Introduction aux données omiques et matrices de géotypes (GWAS)
- Prédiction génomique (Genomic Selection) via modèles linéaires régularisés (Ridge, Lasso)

- Association genotype-phenotype pour l'identification de marqueurs d'interet

11 / Deploiement de modeles et outils d'aide a la decision (OAD)

- Creation d'applications web interactives avec R Shiny ou Python Streamlit
- Conteneurisation des modeles avec Docker pour la mise en production
- APIfication des algorithmes pour l'integration dans des outils tiers

12 / Analyse de survie et epidemiologie vegetale

- Courbes de Kaplan-Meier pour l'etude de la resistance aux maladies
- Modeles de Cox pour analyser les facteurs de risque phytosanitaires
- Simulation de la dispersion des pathogenes dans le paysage

13 / Ethique de la donnee et interpretation causale

- Distinction entre correlation et causalite dans les systemes biologiques complexes
- Biais algorithmiques et equite dans les modeles de recommandation agricole
- Souverainete des donnees agricoles et aspects reglementaires (GDPR)

14 / Projet integrateur : Du Dataset a la Valorisation

- Hackathon sur un jeu de donnees reel (ex: prediction qualite ble ou detection maladies)
- Pipeline complet : Nettoyage -> Modelisation -> Validation -> Visualisation
- Restitution et Storytelling de la donnee pour des decideurs non-experts

Approche pédagogique

- ✓ Support Ecrit et Projection
- ✓ Exposés Interactifs, Podcasts et Vidéos
- ✓ Brainstorming et Jeux de Rôle
- ✓ Mises en Situation pour faciliter l'assimilation
- ✓ Cas Pratiques et Labs inclus pour leur impact opérationnel
- ✓ Test de Validation des Acquis des Connaissances

Prochaines dates programmées

 24 Août au 04 Sep. 2026

 Présentiel -

 19 au 30 Oct. 2026

 Présentiel -

 Autres dates possibles sur demande. Contactez-nous pour organiser une session intra-entreprise.

Réservation & Renseignements

 **Téléphone** : +212 522 247 210

 **Email** : contact@afrique-formation.com

 **Web** : <https://www.afrique-formation.com>