

# Déployer des modèles R dans Databricks (presque) sans douleur

Paul PONCET\*      Gabrielle DEVAUX†

## Résumé

Dans de nombreuses organisations, le langage R est utilisé pour de la modélisation statistique, l'analyse exploratoire ou la construction de prototypes. Lorsque ces modèles doivent être déployés dans des plateformes cloud modernes telles que Databricks, le passage de la théorie à la pratique se transforme souvent en véritable parcours du combattant. Bien que Databricks revendique un support officiel du langage R, les limites apparaissent rapidement dès que l'on s'éloigne des usages standards : gestion complexe des dépendances, gestion des logs en calcul distribué, pipelines mixtes R/Python. Cette présentation propose un retour d'expérience concret sur la migration et le déploiement de packages R internes dans un environnement Databricks orienté vers la production.

À travers quelques exemples, nous explorerons :

- les problèmes rencontrés pour exécuter du R dans Databricks
- la construction et le chargement de packages internes
- comment s'adapter au Runtime Databricks
- transitionner du traditionnel rapport Rmarkdown au notebook Databricks
- les bonnes (et mauvaises) pratiques permettant d'assurer une transition sereine vers un environnement MLOps moderne

Au-delà des aspects techniques, cette présentation invite à une réflexion plus large : comment continuer à valoriser l'écosystème R dans un monde dominé par des infrastructures pensées d'abord pour Python et Spark ? Un témoignage sans filtre, mêlant frustrations, astuces et enseignements utiles pour tous ceux qui doivent « faire tourner du R » dans Databricks.

**Mots-clefs (3 à 5) :** Ingénierie – Data – Package – Industrialisation – Databricks

\* Data Scientist Engie, [paul.poncet@engie.com](mailto:paul.poncet@engie.com)

† Consultante Data Lincoln chez Engie, [gabrielle.devaux@mel.lincoln.fr](mailto:gabrielle.devaux@mel.lincoln.fr)