

# Mise en production de modèles Machine Learning avec PMML dans le secteur bancaire

Pidash, Angelina, angelina.pidash@bnpparibas-pf.com, BNP Paribas Personal Finance, Bordeaux

*Résumé : Le secteur bancaire est un terrain idéal pour différentes applications de Machine Learning, mais très souvent, les infrastructures informatiques actuellement en place ne sont pas particulièrement flexibles et innovantes pour pouvoir les mettre en production. Dans ce cadre, nous allons voir comment nous avons mis en place une « traduction » de nos modèles Machine Learning pour être compris par notre système en production et comment l'appliquer au travers d'un exemple.*

Mots clés : *#machinelearning #pmml #scores #banque #production*

## 1. Introduction

Au sein de notre équipe, nous testons les différentes approches Machine Learning pour développer des modèles de risque de crédit ou de détection de fraude. Un vrai challenge pour nous consiste à pouvoir les mettre en production dans les différents environnements de production qui varient entre les filiales et pays. Pour résoudre ce problème, nous nous sommes tournés vers PMML (*Predictive Model Markup Language*), un format basé sur XML qui permet d'exporter les modèles Machine Learning entraînés puis de les réutiliser dans différents environnements autres que celui d'entraînement.

## 2. Méthodologie

- Machine Learning: méthodes ensemblistes comme Random forest ou Gradient boosting.
- Transformation des modèles Machine Learning en PMML en utilisant le package *sklearn2pmml* de Python.
- Utilisation d'un fichier PMML avec Python, Spark ou autre.

## 3. Originalité / perspective

PMML permet d'automatiser plusieurs étapes et optimiser le processus de mise en production des modèles de score. En perspective, nous envisageons d'intégrer la partie « Feature engineering » dans le pipeline complet de PMML et tester d'autres types de modèles Machine Learning.