

# Rapport : M2-BigData-MLIB-SupervisedLearning

## Support Vecteur Machine (SVM)

Le meilleur modèle de Support Vecteur Machine est celui qui donne une ROC (AUC) de 0.9001. Ce résultat est obtenu en optimisant les hyperparamètres suivants :

**maxIter**: 10

**regParam**: 0.1

En termes de performance, le modèle présente un F1-Score de 0.8399 et un Recall de 0.8472.

## Random Forest

Le meilleur modèle Random Forest est celui qui donne une ROC (AUC) de 0.9103. Ce résultat est obtenu en optimisant les hyperparamètres suivants :

**numTrees**: 30

**maxDepth**: 15

En termes de performance, le modèle présente un F1-Score de 0.8495 et un Recall de 0.8569.

## Naive Bayes

Le modèle Naive Bayes optimisé ne performe pas aussi bien que les modèles de classification binaire. Il présente une ROC (AUC) de 0.3585, un F1-Score de 0.7457, et un Recall de 0.7837. Ces résultats indiquent que le modèle Naive Bayes ne parvient pas à capturer de manière optimale les relations dans les données.

Ce résultat est obtenu en optimisant l'hyperparamètre suivant : **smoothing**: 0.1

- 
- ➔ Pour chaque modèle, j'ai utilisé **ParamGridBuilder** pour définir une grille d'hyperparamètres à tester lors de la recherche d'optimisation. Cela permet de sélectionner les meilleurs hyperparamètres pour maximiser les performances du modèle. Chaque grille est spécifique aux paramètres du modèle correspondant, par exemple, **maxIter** et **regParam** pour SVM, et **numTrees**, **maxDepth** pour Random Forest et **smoothing** pour Naive Bayes
  - ➔ Les résultats montrent que le modèle Random Forest surpasse les autres modèles en termes de performance

```
Optimized SVM Area under ROC curve: 0.9000836055585126
Optimized SVM F1-Score: 0.8398780591912192
Optimized SVM Recall: 0.8471858134155743
Optimized Random Forest Area under ROC curve: 0.9103137553430818
Optimized Random Forest F1-Score: 0.8495389998101303
Optimized Random Forest Recall: 0.8569005397070162
Optimized Naive Bayes Area under ROC curve: 0.35845595740315533
Optimized Naive Bayes F1-Score: 0.7456501360021339
Optimized Naive Bayes Recall: 0.7836545875096376
```