MAL Practical Evaluation 2025

Adam DJAROUD ENSIIE Clément LÉVÊQUE ENSIIE

Décembre 2025

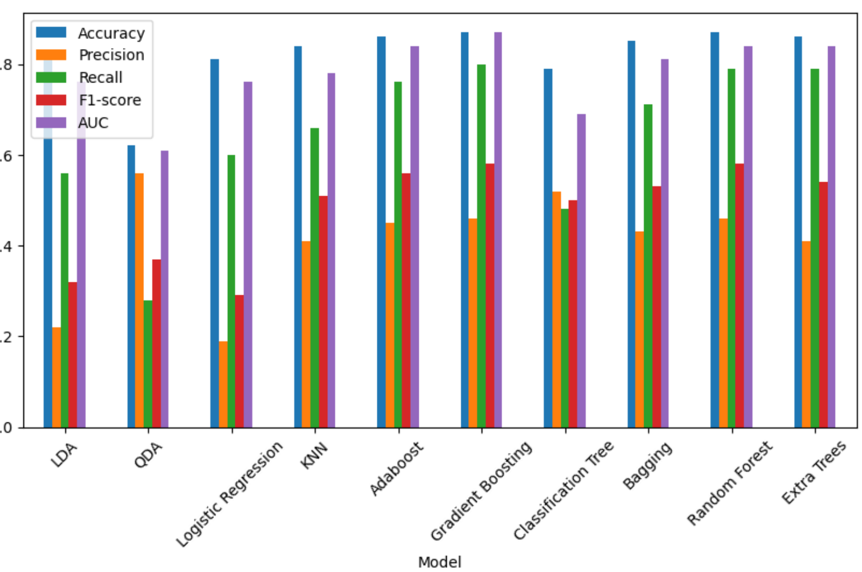
Question 1

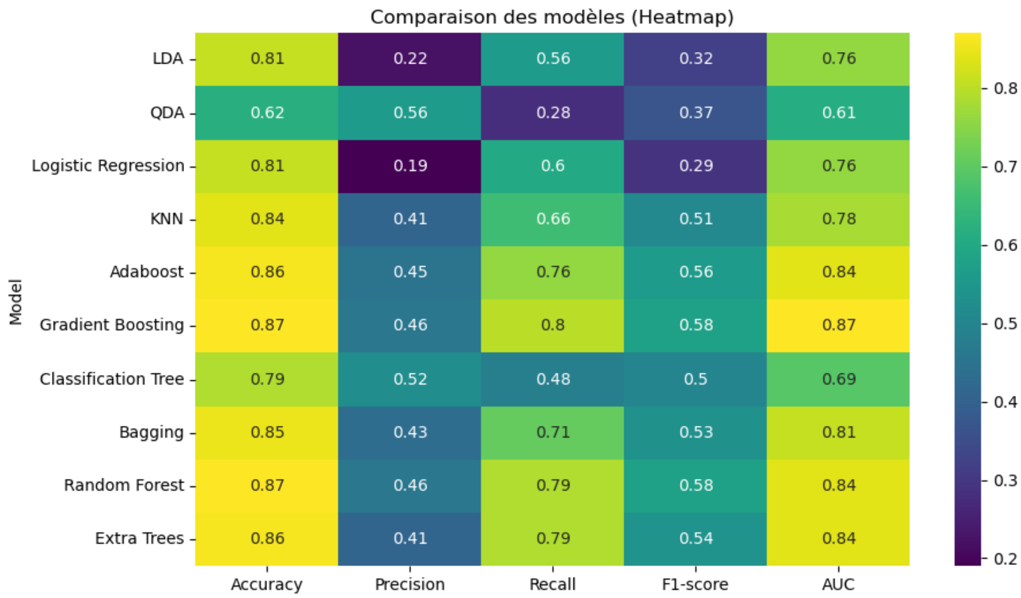
Introduction

To answer this question, we tested all of the models studied in the MAL course. We know that in general, the most efficient models for classification are random forest and gradient boosting, so we expect to see this in the results.

Models

To use `scikit-learn`, we first needed to encode the categorical variables with the `OneHotEncoder` from `scikit-learn`. Then we separated the data into train and test, with 30% test.
After training and testing the different models on the data, we plotted the results :





As we expected, `Gradient Boosting` and `Random Forest` are the models that stand out from the group. We also see some outliers, like the `Precision` from `QDA` which is the best `Precision`. However its `Accuracy` is close to 50% which indicates that it is close to random.
Overall, the metrics are all consistent with each other, therefore to pick the best model, we can just pick the one with the best `AUC`, which is the `Gradient Boosting` model.

So we would advice the director of the mobile phone operator to pick the `Gradient Boosting` model for predicting the churn rate of its company.

Question 2

Considering the variable `Gender` as a sensitive attribute, we will assess the fairness, independence, and sufficiency of the `Gradient Boosting` model. The analysis of the three fundamental criteria (Independence, Separation, and Sufficiency) reveals how the model behaves between the Male and Female groups.

- Analysis of the Independence Criterion (Demographic Parity)
  The Independence criterion is evaluated by comparing the predicted churn rate ($\mathbb{P}(\hat{Y} = 1|Gender)$) between groups. Our results show a slight violation of the Independence criterion with an absolute difference of 0.0517. The model predicts churn for 14.38% of females compared to only 9.21% of males. This violation indicates that the `Gradient Boosting` model targets female customers significantly more often than male customers for retention campaigns. While this difference may reflect a real correlation between gender and churn rate in the data (i.e., the actual `Churn` rate is higher among Women), it raises concerns about targeting bias.

- Analysis of the Separation Criterion (Equal Opportunity): The Separation criterion, which is often the most relevant for retention, requires equality of Recall (`Recall`): the detection rate of customers who actually churn must be the same for all groups. The results show excellent compliance with the Separation criterion, with an absolute difference of only 0.0022.

  Recall for Females: 45.85 %

  Recall for Males: 45.63 %

  This proximity is crucial. It ensures that the `GBM` model is equally effective at identifying at risk customers, regardless of their gender. The Director can be assured that the model does not exhibit bias in its effectiveness to save customers, ensuring equality between men and women in benefiting from a retention offer.

- Analysis of the Sufficiency Criterion (Predictive Equality): The Sufficiency criterion is evaluated by comparing Precision (`Precision`), which measures the reliability of churn prediction for each group. The results indicate that this criterion is `CLOSELY MET`, with a notable difference of 6.58%.

  Precision for Females: 83.01 %

  Precision for Males: 76.42 %

  This means that churn predictions for Females are more reliable. Targeting Males will result in a higher rate of false positives (customers incorrectly targeted).

Conclusion

The results show that the model maintains excellent fairness in its detection capability, meaning there is no bias in error when identifying customers to save.

However, since independence is violated, femeales are over targeted. Sufficiency is CLOSELY MET, which means that targeting males generates more false positives. Both pieces of information should be taken into account by the Director to prevent churn.