

PRÉDICTION DU CHURN EN ASSURANCE AUTOMOBILE SUR AZURE CLOUD

Par El Harrar Ethan

Encadré par Mr. Thibault Malherbe



INTRODUCTION ET CONTEXTE

- **Rétention Client** : Fidéliser coûte moins cher qu'acquérir de nouveaux clients et augmente la valeur à vie.
- **Impact du Churn** : Le churn réduit les revenus, augmente les coûts d'acquisition et affecte la rentabilité.
- L'objectif de ce projet sera :
 - Utiliser Azure Cloud (Blob Storage, Databricks, AutoML).
 - Créer un modèle de Machine Learning capable de prédire la probabilité de départ des clients.
 - Réaliser le projet en respectant le budget prédefini. (< 20 euros)
 - Comparer les résultats obtenus avec d'autres modèles développés via AutoML.



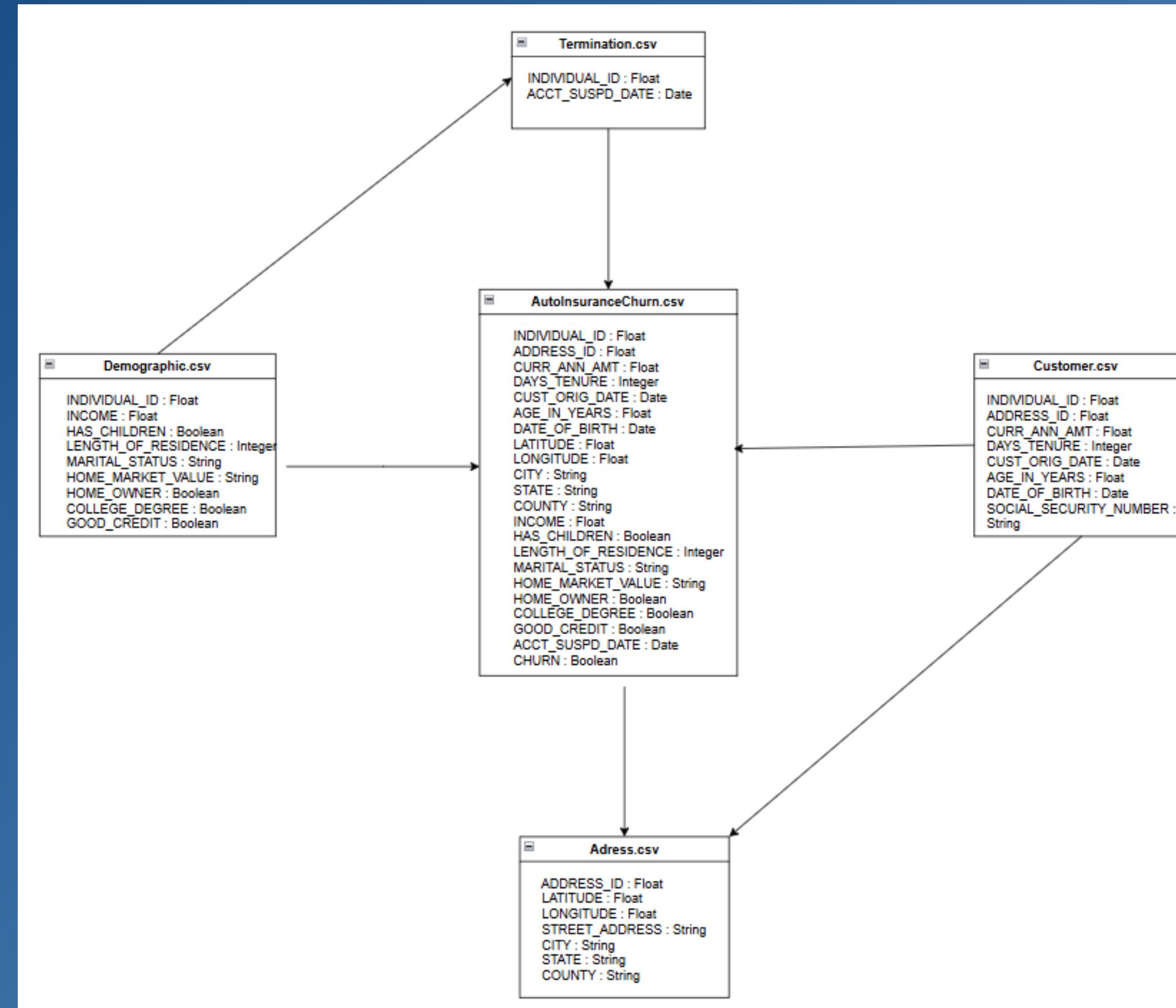
DATASET

The screenshot shows a dataset page from a platform. At the top left is a profile picture of a woman and the text "MERISHNA SINGH SUWAL · UPDATED 2 YEARS AGO". To the right are navigation icons for back, forward, and search, followed by a page number "12", a "Code" button, a "Download" button with a downward arrow icon, and a more options menu. The main title "Auto Insurance churn analysis dataset" is displayed prominently in large, bold, dark font. Below the title is a subtitle: "Synthetic dataset that includes information about auto insurance customers." To the right of the title is a light gray callout box containing the text "CUSTOMER CHURN" above a small illustration of a fish tank with three fish, one of which is jumping out, symbolizing customer churn.

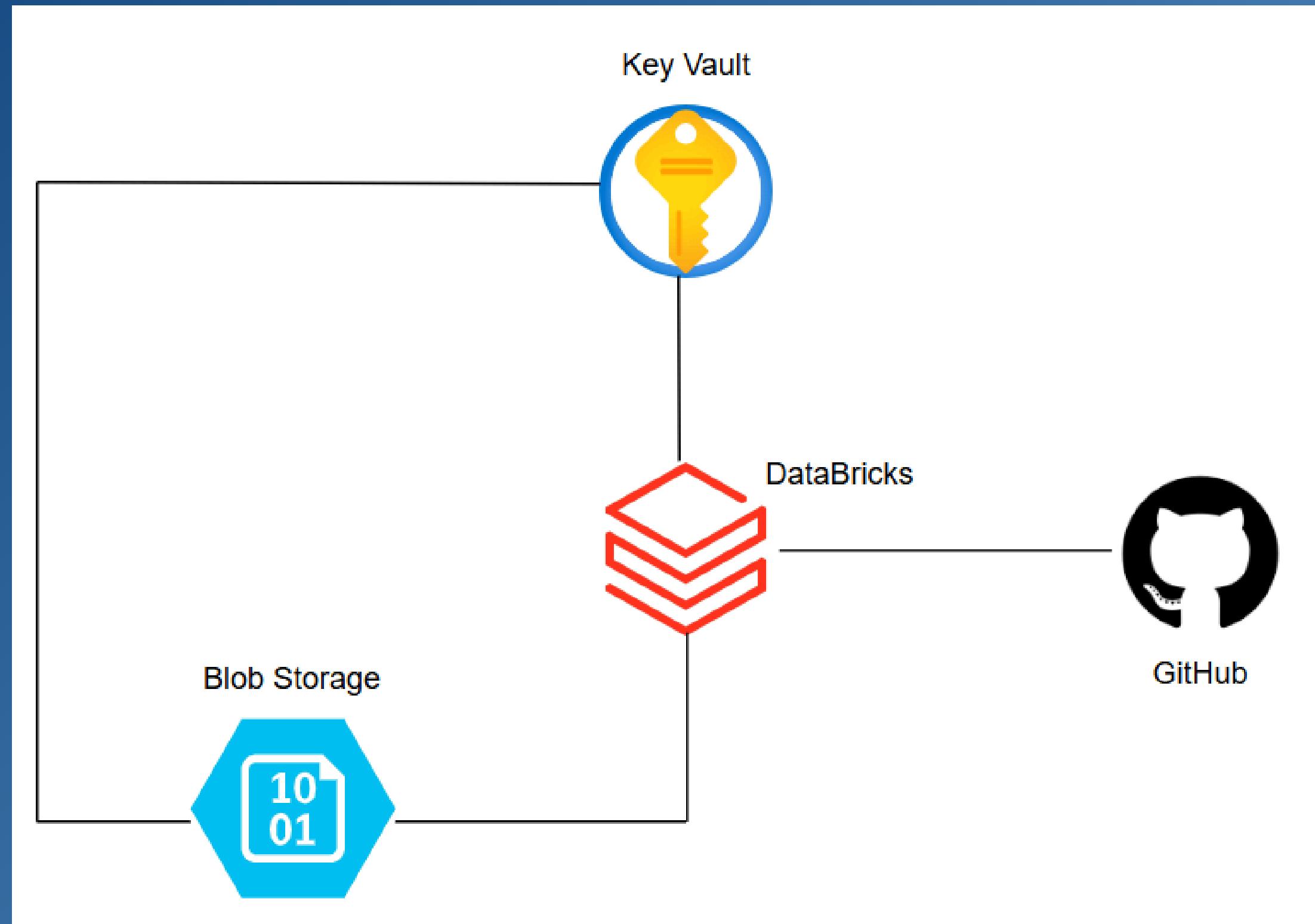
À propos du dataset:

- **address.csv:** données d'adresse (116mb, +1mio lignes)
- **customer.csv:** données de client (197mb, +1mio lignes)
- **demographic.csv:** données géographique (120mb, +1mio lignes)
- **termination.csv:** données sur la résiliation du client (6mb, 269K lignes)
- **autoinsurance_churn.csv:** données fusionnées sur la perte de clients générées par ce notebook. (273mb, +1 mio lignes)

ARCHITECTURE DES TABLES



ARCHITECTURE APPLICATIVE



NOTEBOOK'S WORKSPACE

Workspace > Users >

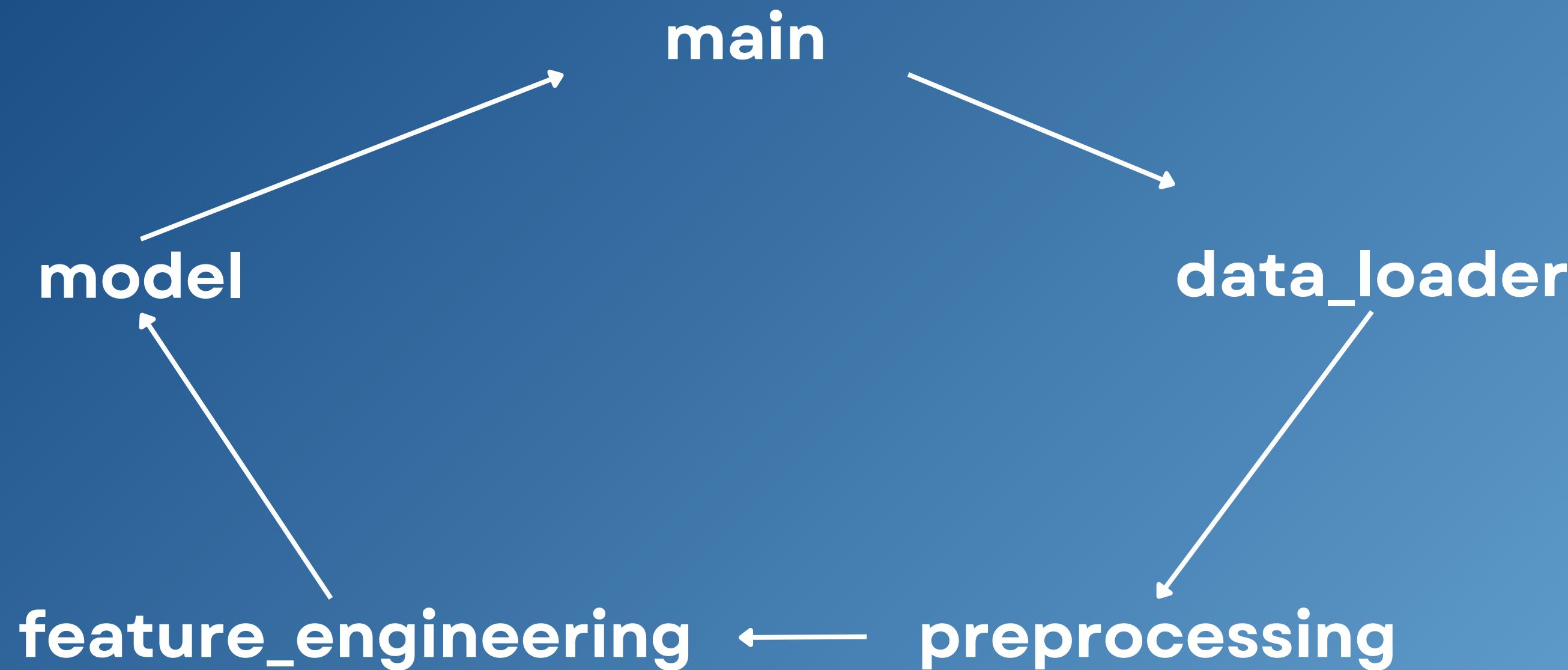
eelharrar@myges.fr ☆

⋮ Partager Créeer ⋮

| Nom ⌗↑ | Type | Propriétaire | Créé à | ⋮ |
|------------------------------|----------|--------------|----------------------|---|
| 📁 databricks_automl | Dossier | Ethan HARRAR | 11 févr. 2025, 18:33 | ⋮ |
| 📁 data_loader | Notebook | Ethan HARRAR | 10 févr. 2025, 16:55 | ⋮ |
| 📁 feature_engineering | Notebook | Ethan HARRAR | 10 févr. 2025, 16:57 | ⋮ |
| 📁 Main | Notebook | Ethan HARRAR | 10 janv. 2025, 16:21 | ⋮ |
| 📁 model | Notebook | Ethan HARRAR | 10 févr. 2025, 16:58 | ⋮ |
| 📁 modified_tables_for_automl | Notebook | Ethan HARRAR | 11 févr. 2025, 19:04 | ⋮ |
| 📁 preprocessing | Notebook | Ethan HARRAR | 10 févr. 2025, 16:58 | ⋮ |



CHEMINEMENT



PRETRAITEMENT DES DONNÉES

-> **Preprocessing**

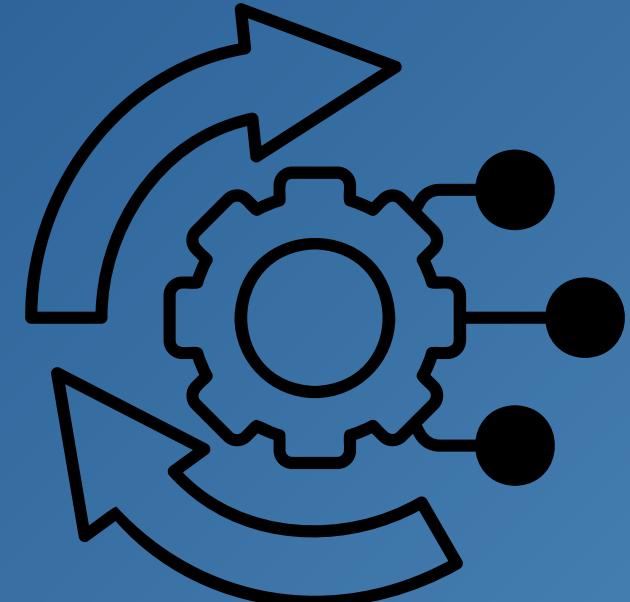
- 1) Lecture et Fusion des données** (dataframe unifié)
- 2) Nettoyage des données** (conversion, suppression de duplication)
- 3) Transformation Spécifique**
- 4) Imputation des Valeurs Manquantes** (replace nulles par moyenne)
- 5) Sauvegarde du résultat** (preprocessed_table)

PRETRAITEMENT DES DONNÉES

-> **feature_engineering**



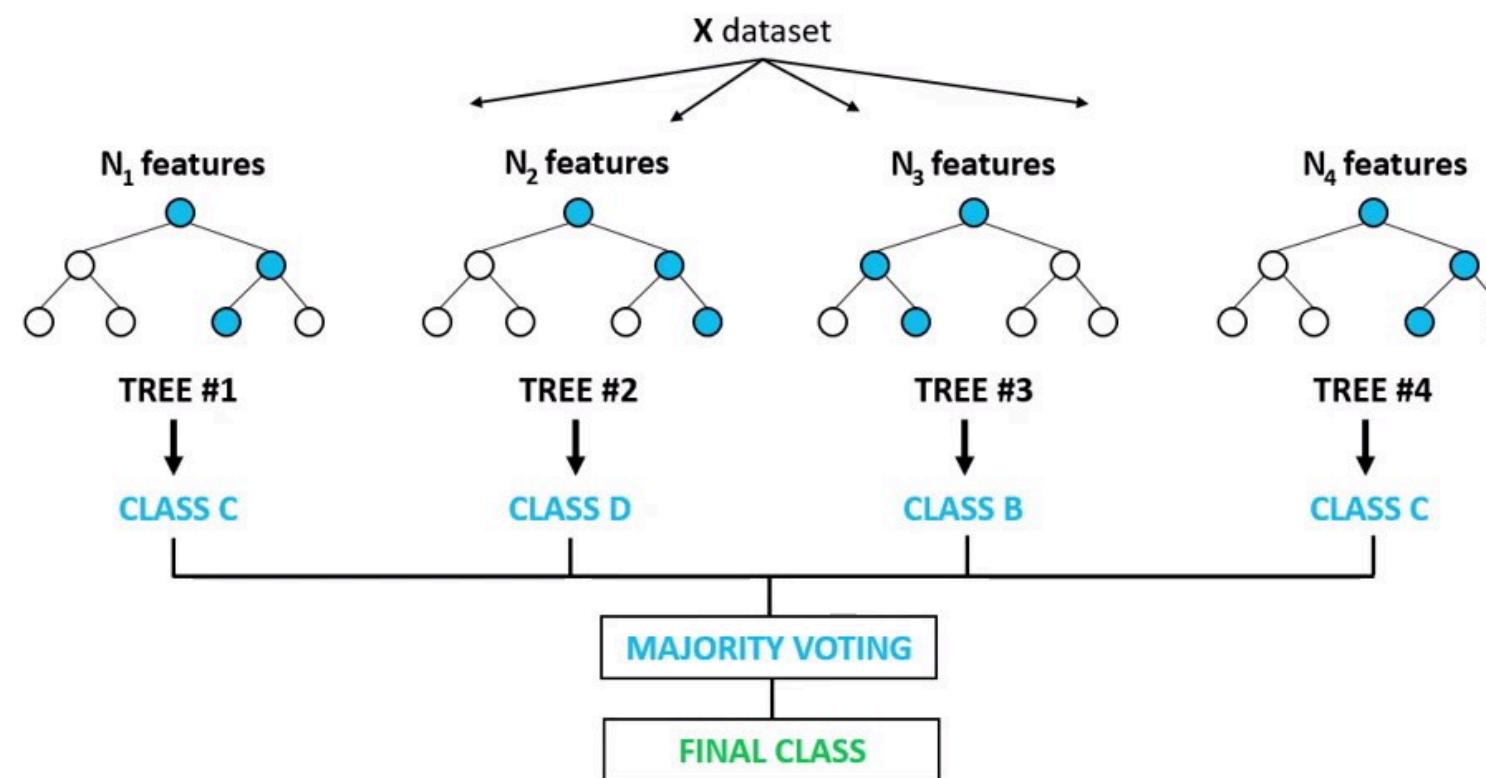
preprocessed_table



features_table
Spark Table

MODEL

Random Forest Classifier



numTrees : 100

seed : 42

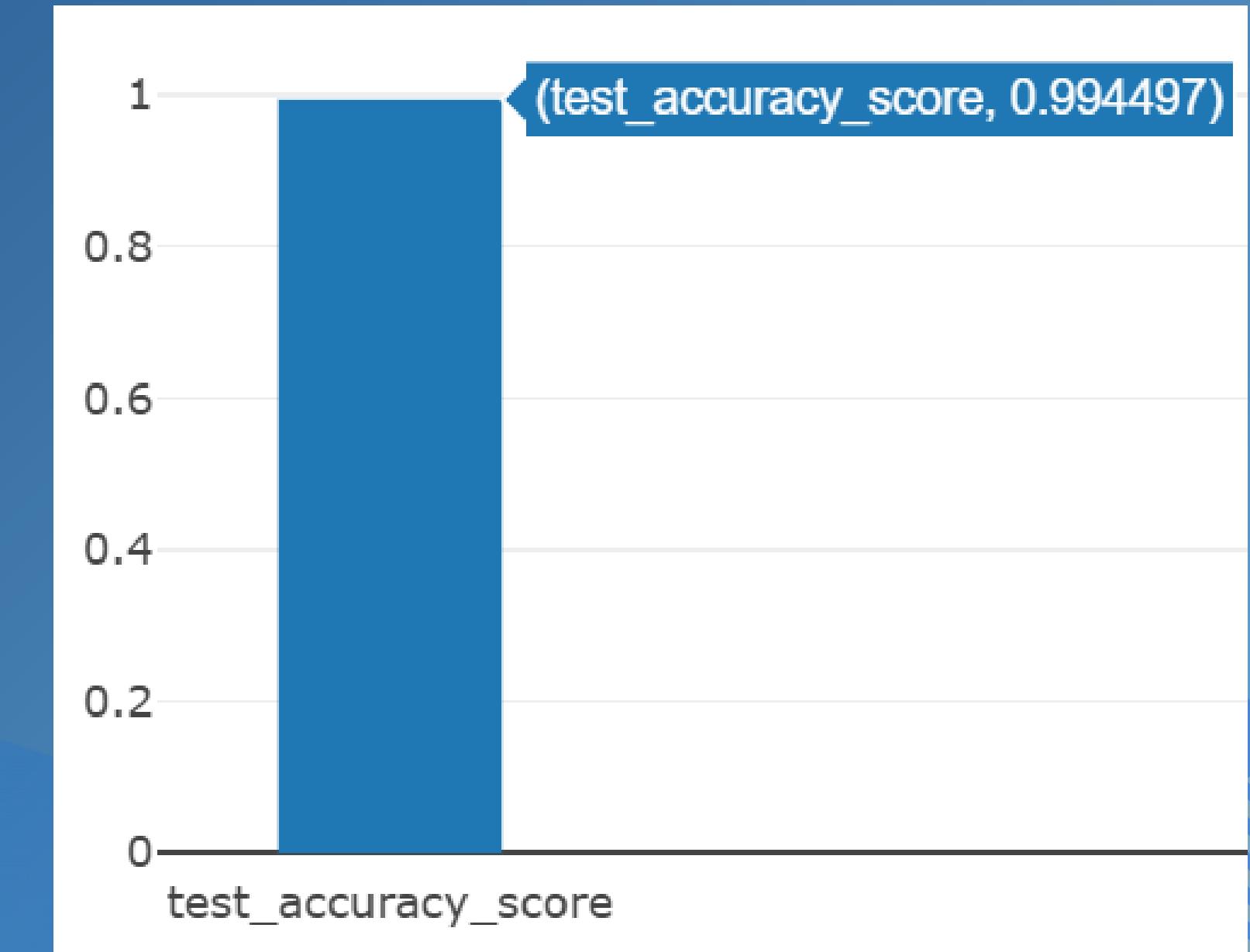
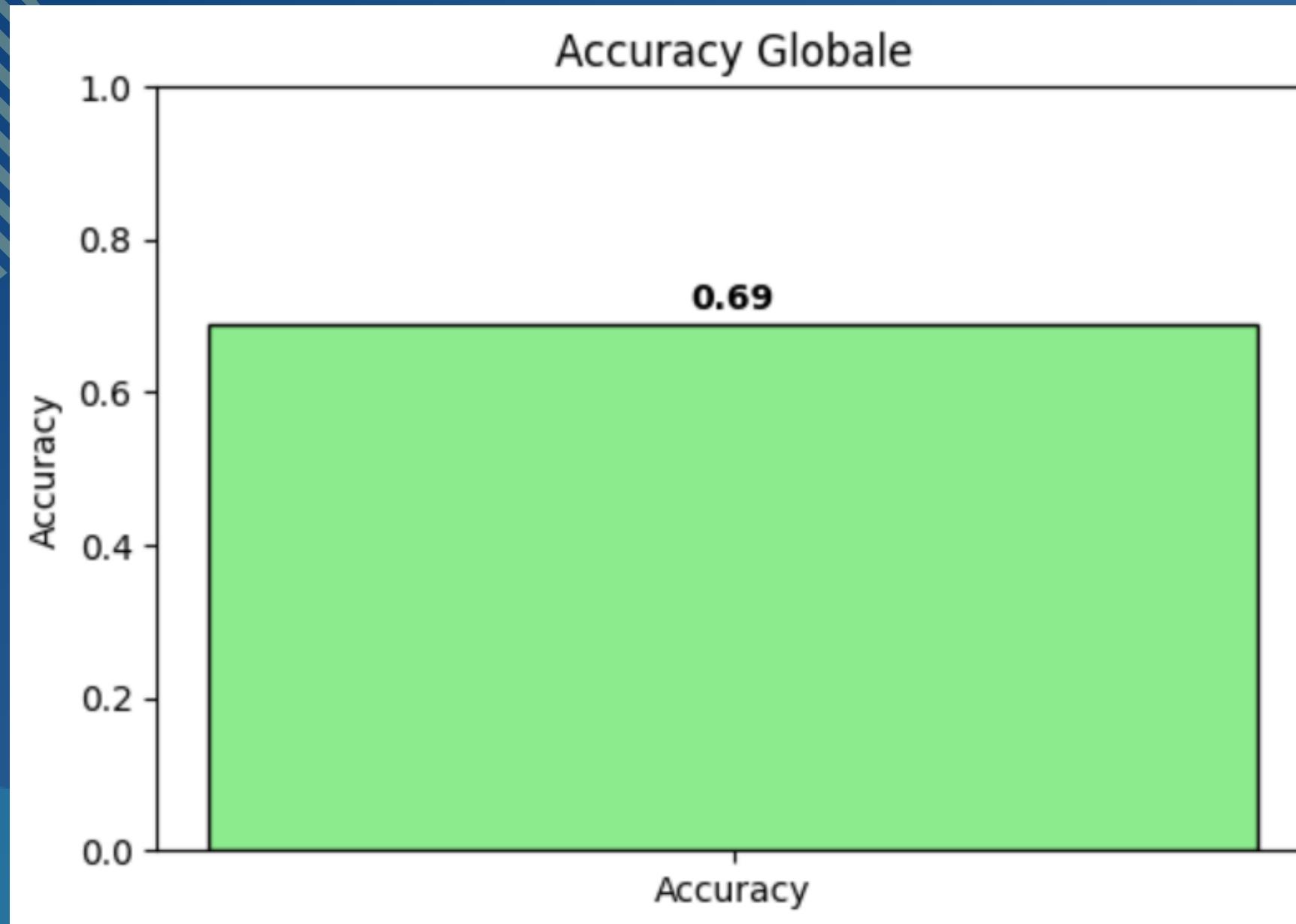
maxBins : 100

AUTOML

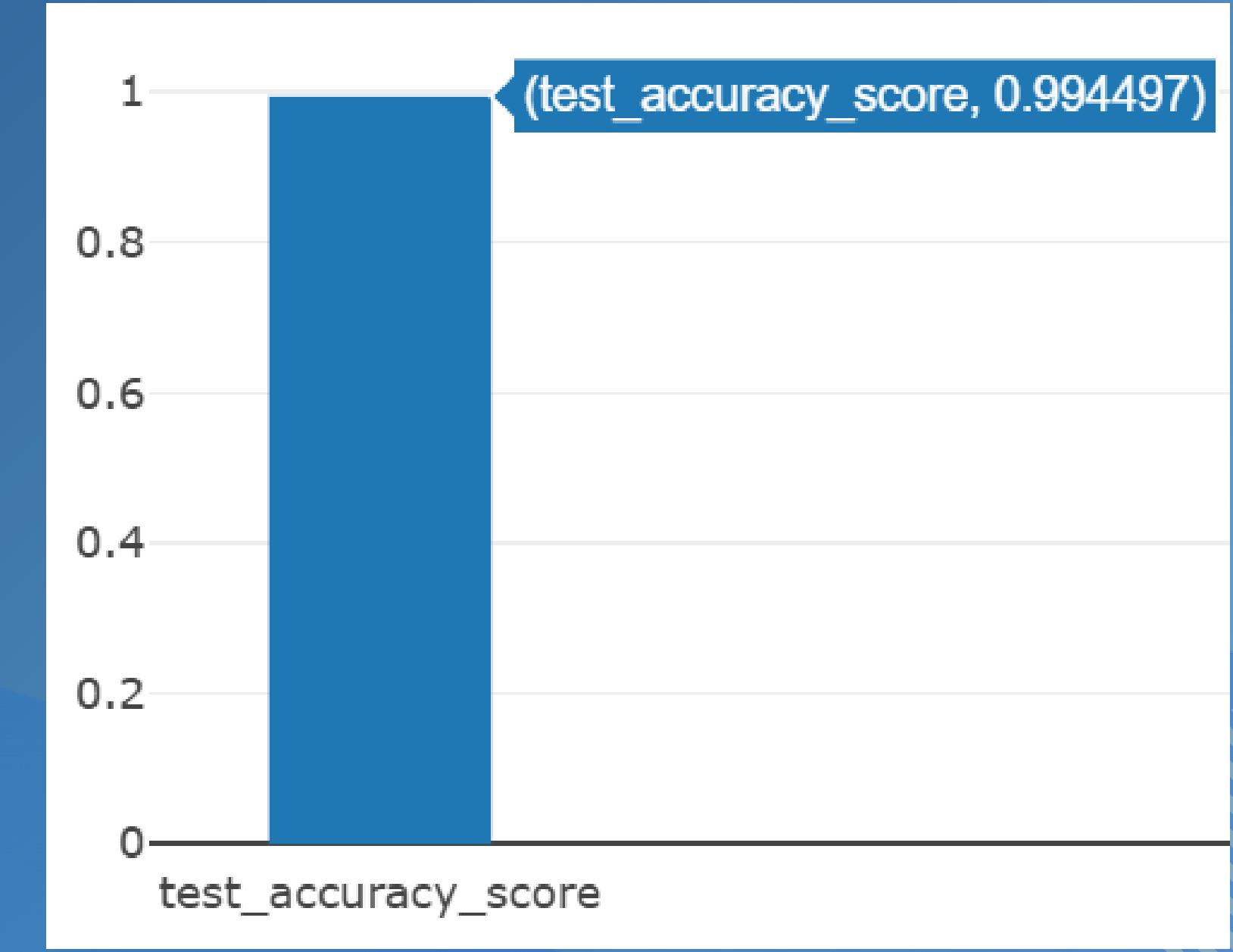
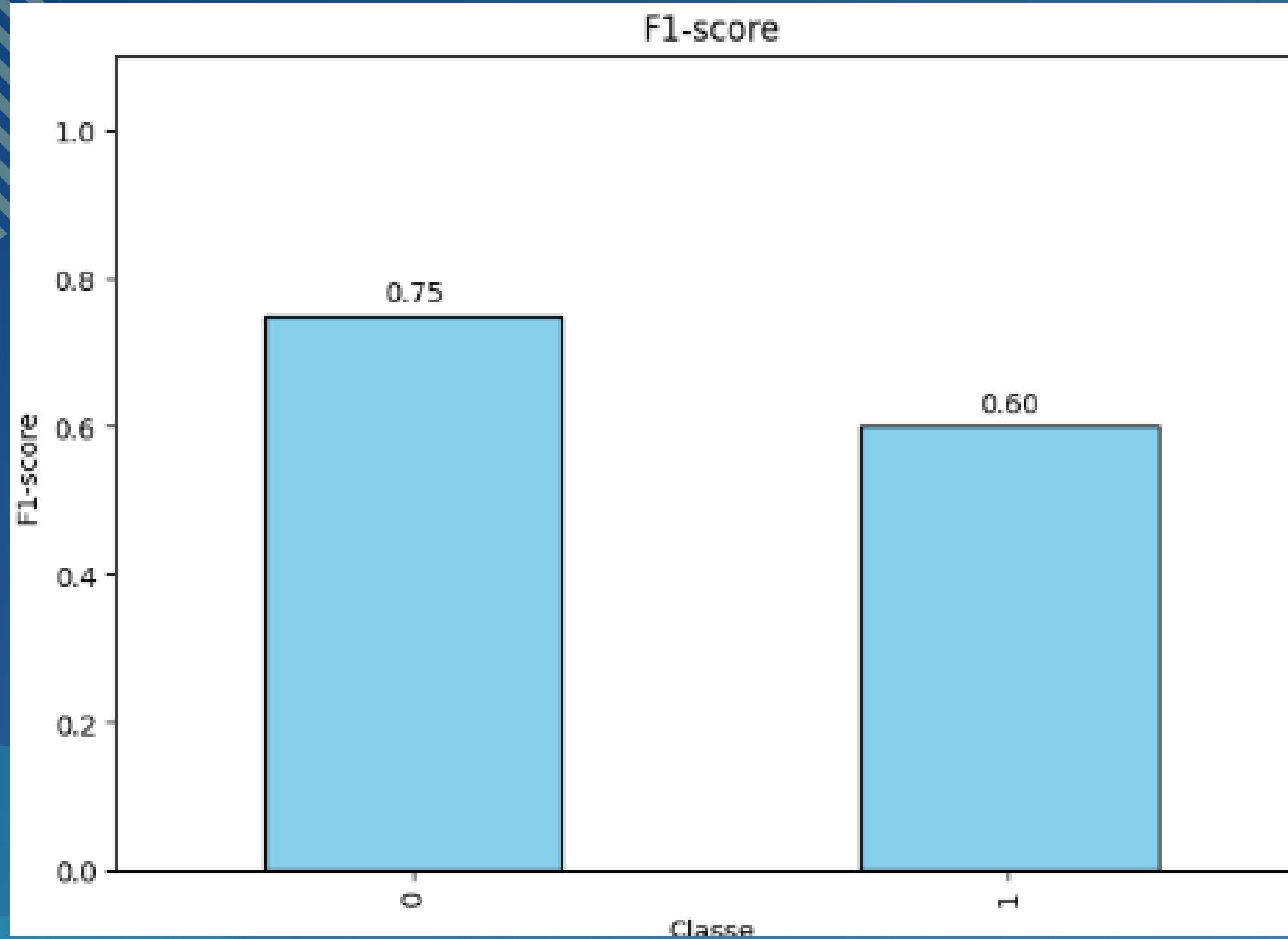
Table Graphique Évaluation Preview

| | Run Name | Created | Source | Models | Metrics | Tags |
|--------------------------|----------------------|----------------------------------|--------|---|--|--------------------------|
| <input type="checkbox"/> | gaudy-slug-778 | ✓ Il y a 19 minutes | - |  sklearn | val_accuracy_score  1 | lightgbm_classifier |
| <input type="checkbox"/> | intelligent-snail-53 | ✓ Il y a 19 minutes | - |  sklearn | 1 | xgboost_classifier |
| <input type="checkbox"/> | marvelous-quail-442 | ✓ Il y a 31 minutes | - |  sklearn | 1 | random_forest_classifier |
| <input type="checkbox"/> | brawny-carp-500 | ✓ Il y a 42 minutes | - |  sklearn | 0.9939687994878533 | lightgbm_classifier |

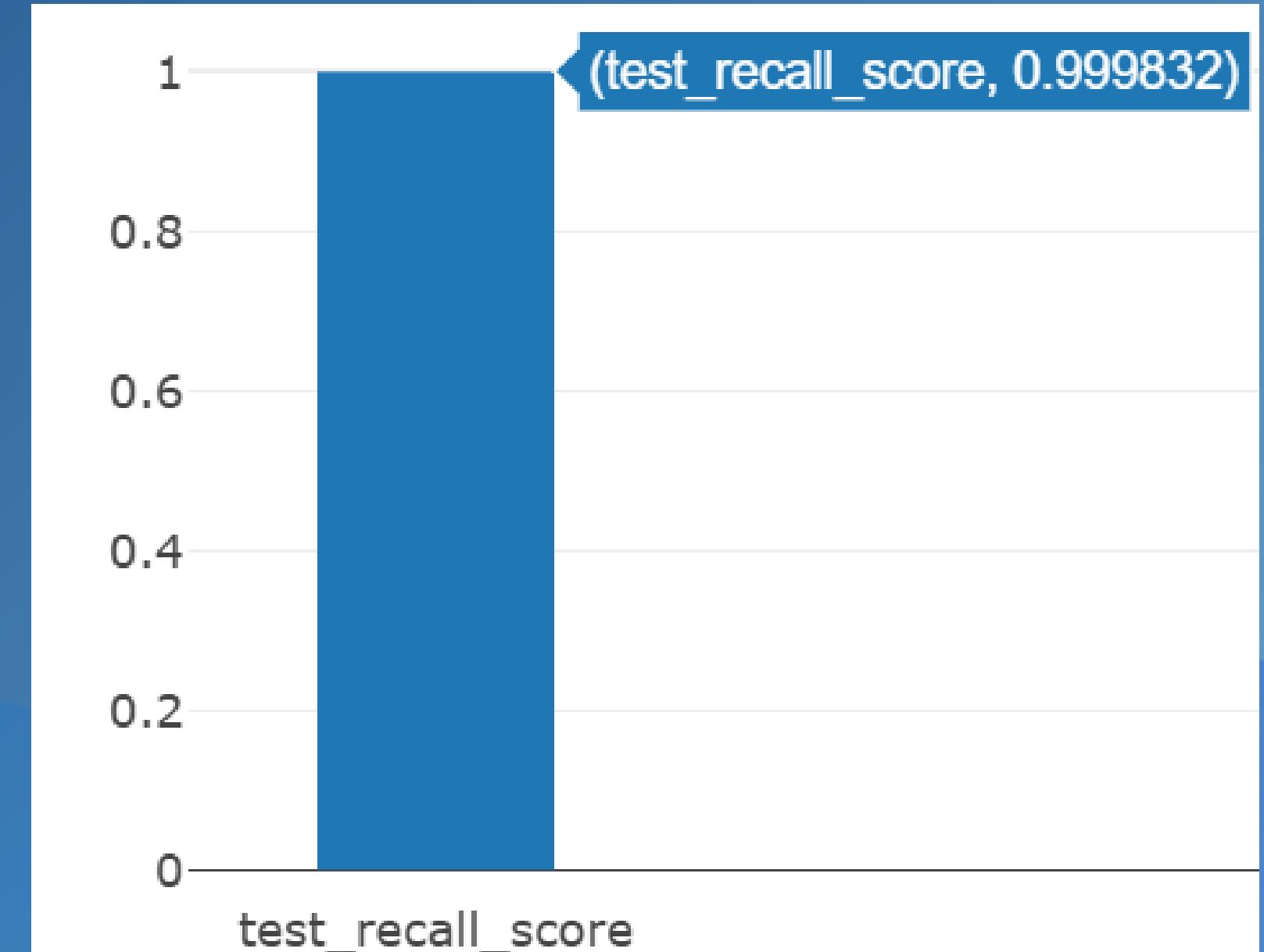
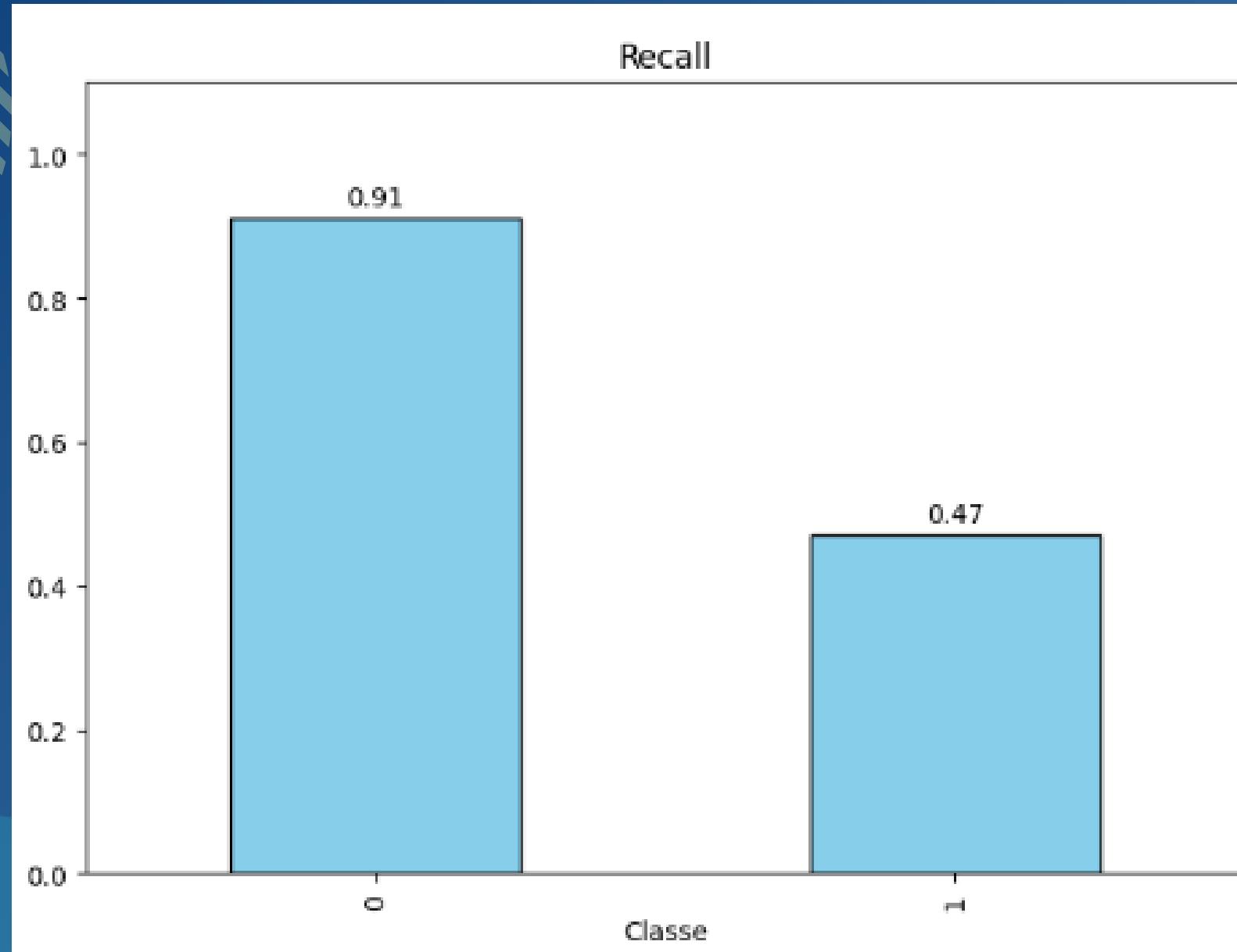
RANDOMFOREST VS AUTOML



RANDOMFOREST VS AUTOML

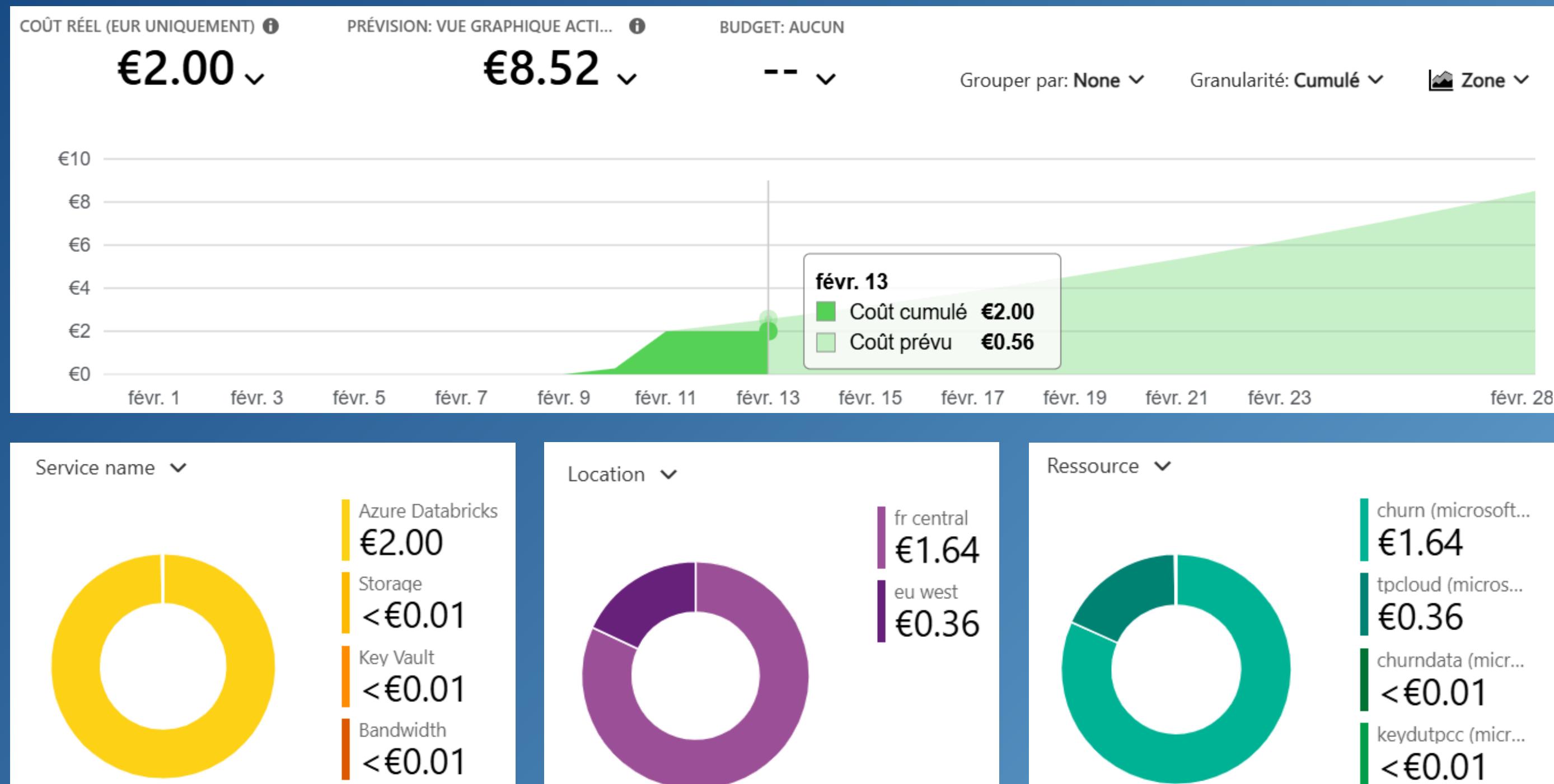


RANDOMFOREST VS AUTOML



BUDGET

A-t-on dépassé le budget de moins de 20€ ?





CONCLUSION

- Azure Cloud ?
- Créer un modèle de ML ?
- Budget ?
- Model vs AutoML ?

AXE D'AMELIORATION

- 1) Amélioration du dataset**
- 2) Réentraînement du modèle**
- 3) Adaptation du modèle dans un contexte de banque et téléphonie**

MERCI POUR TOUS!