

CI/CD pour le Machine Learning GitHub Actions + CML

Saha Merlin

Repository GitHub: <https://github.com/devsahamerlin/mlops-churn-cml-dvc-cicd>

Automatisation MLOps côté GitHub :

- exécuter l'entraînement du modèle à chaque git push sur la main et chaque PR sur la main et la dev,
- générer les métriques (metrics.txt) et la matrice de confusion,
- publier automatiquement un rapport lisible en commentaire sur GitHub grâce à CML.

Vue d'ensemble

1. Vous avez un projet ML classique (dataset.csv, script.py).
 2. Vous poussez ce projet sur GitHub.
 3. GitHub Actions lance automatiquement l'entraînement (CI).
 4. Les perfs du modèle (scores, F1-score, matrices de confusion) sont publiées en commentaire.
C'est utile pour un projet de data science en équipe :
- visibilité instantanée de la qualité du modèle,
 - traçabilité des changements,
 - boucle d'amélioration continue.

Rôle de chaque fichier

- **dataset.csv**: données d'entraînement/test utilisées par le script.
- **script.py**:
 - lit dataset.csv,
 - prépare les features (imputations, encodage, équilibrage),
 - entraîne un modèle (ex. RandomForestClassifier),
 - calcule des métriques (F1, précision, rappel, etc.),
 - génère une figure conf_matrix.png,
 - écrit un résumé clair dans metrics.txt.
- **requirements.txt**: dépendances Python à installer dans la CI.

- [.github/workflows/cml-churn.yaml](#): pipeline GitHub Actions + CML.

Local Environment Setup

```
python3 -m venv .venv
source .venv/bin/activate
pip install --upgrade pip
pip install -r requirements.txt
```

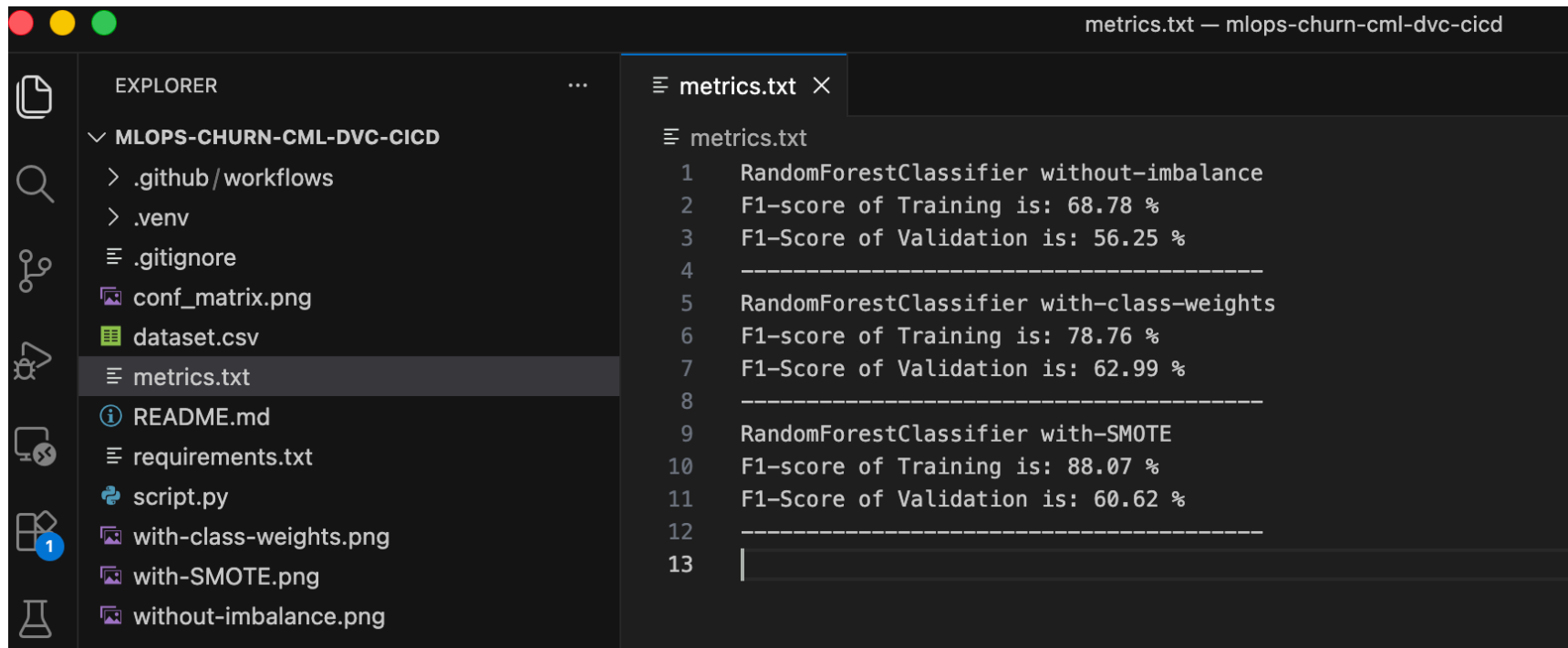
Local test

```
python3 script.py
```

Results Local

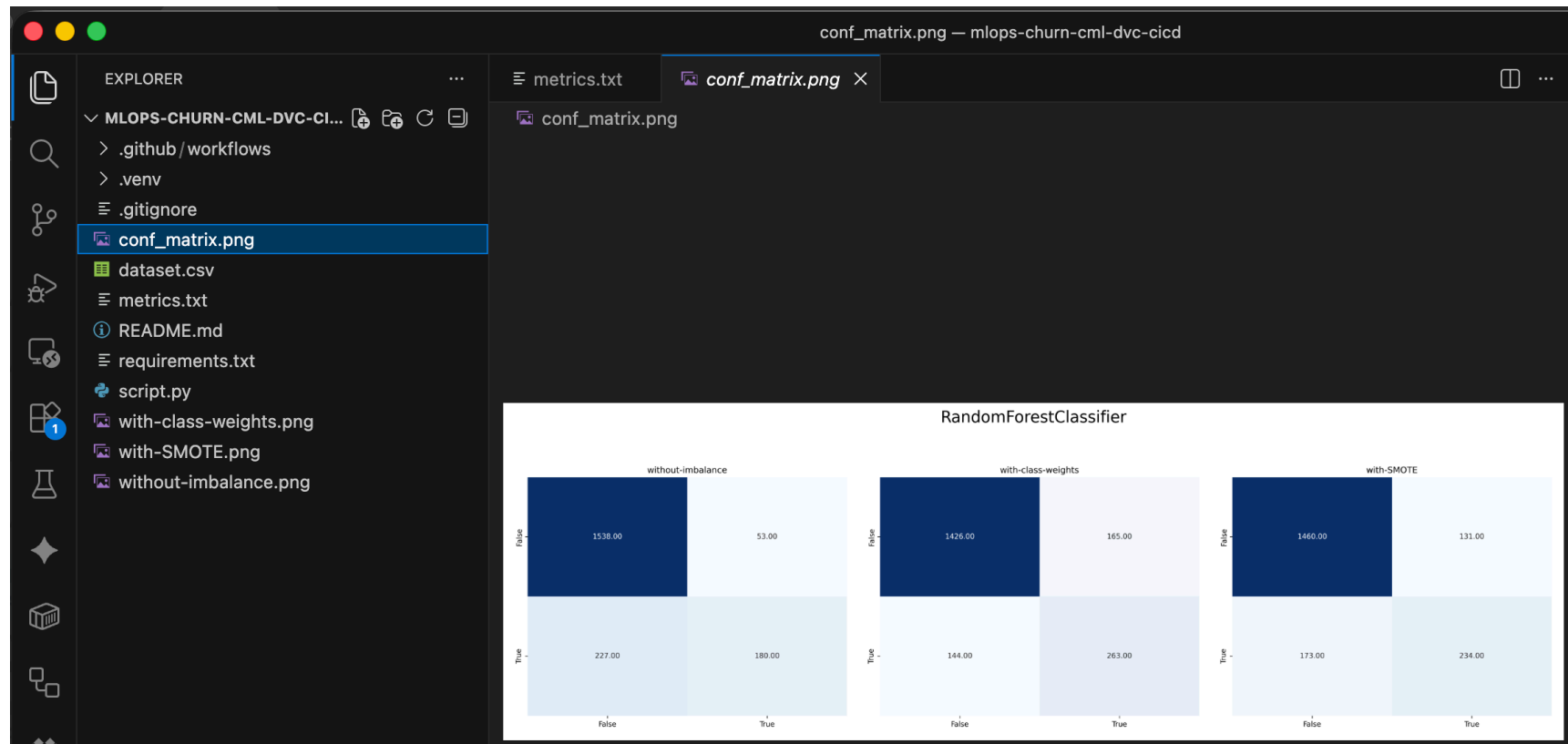
Metrics

Aperçu des métriques de performance (Accuracy, F1-score, ROC-AUC) générées lors de l'exécution locale du script.



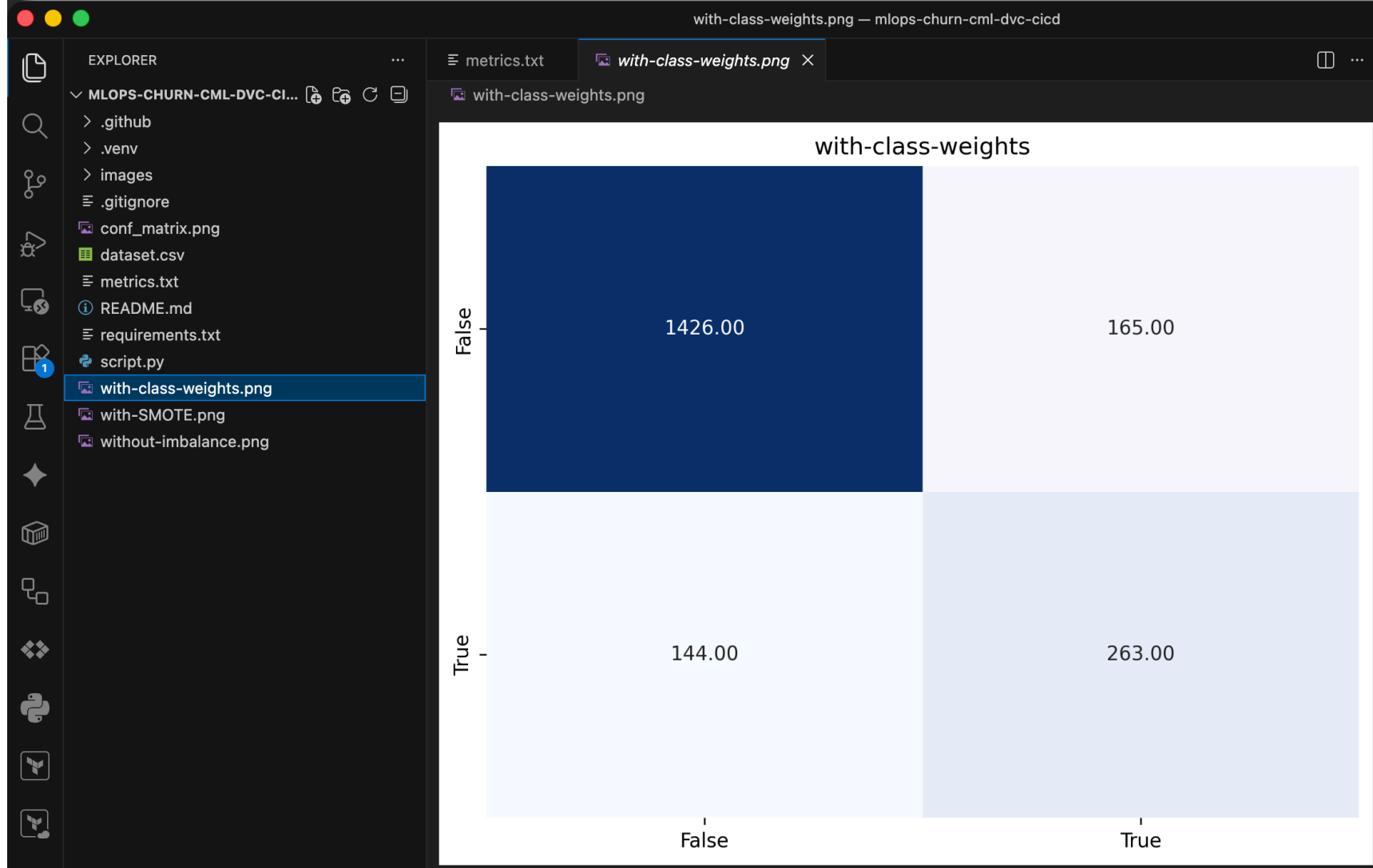
Confusion Matrix

Matrice de confusion montrant la capacité du modèle à bien classer les clients (churn vs non-churn).



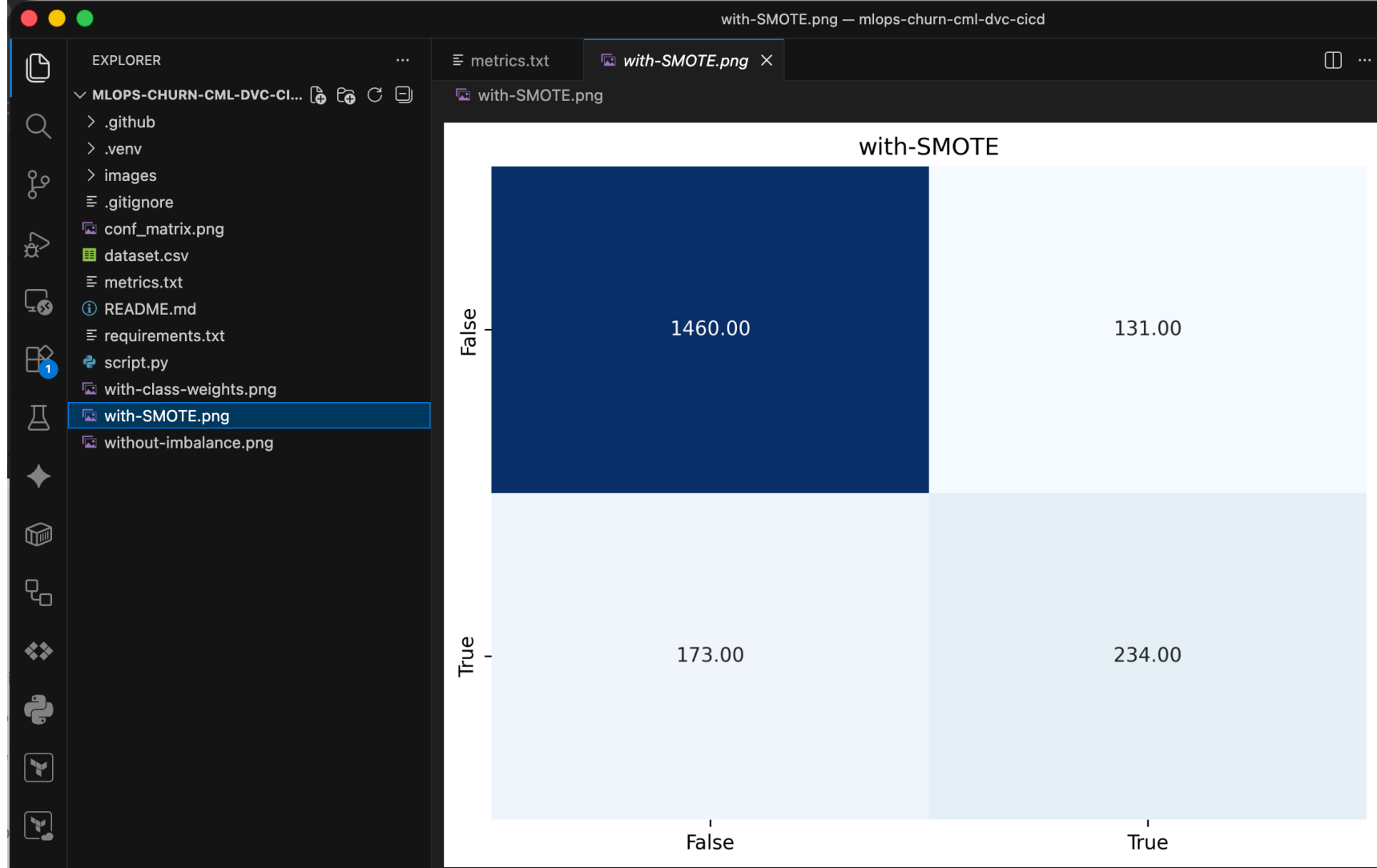
With Class Weights

Résultats obtenus en utilisant des poids de classe pour compenser le déséquilibre des données.



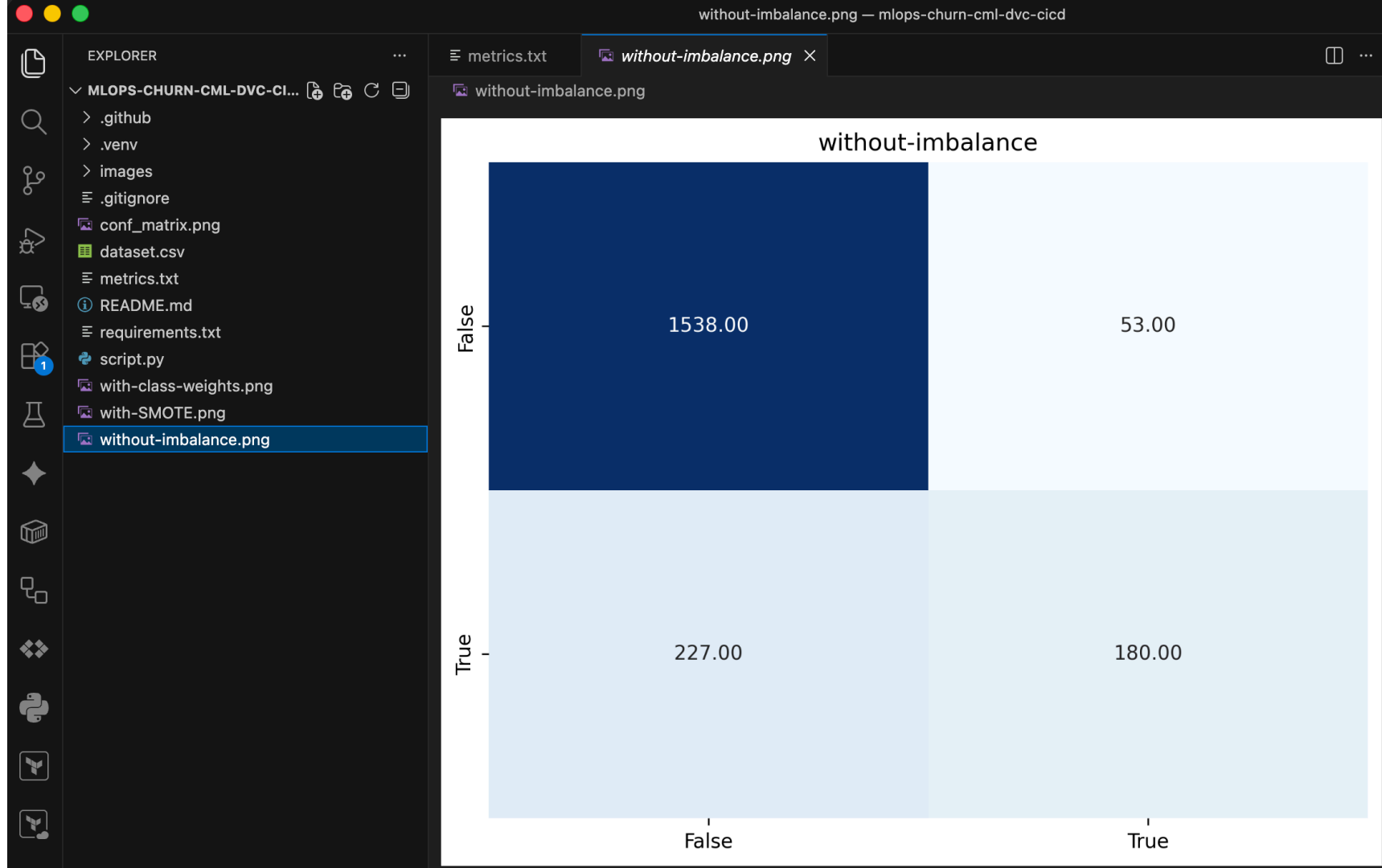
With SMOTE

Résultats après application de SMOTE (Synthetic Minority Over-sampling Technique) pour générer des exemples synthétiques de la classe minoritaire.



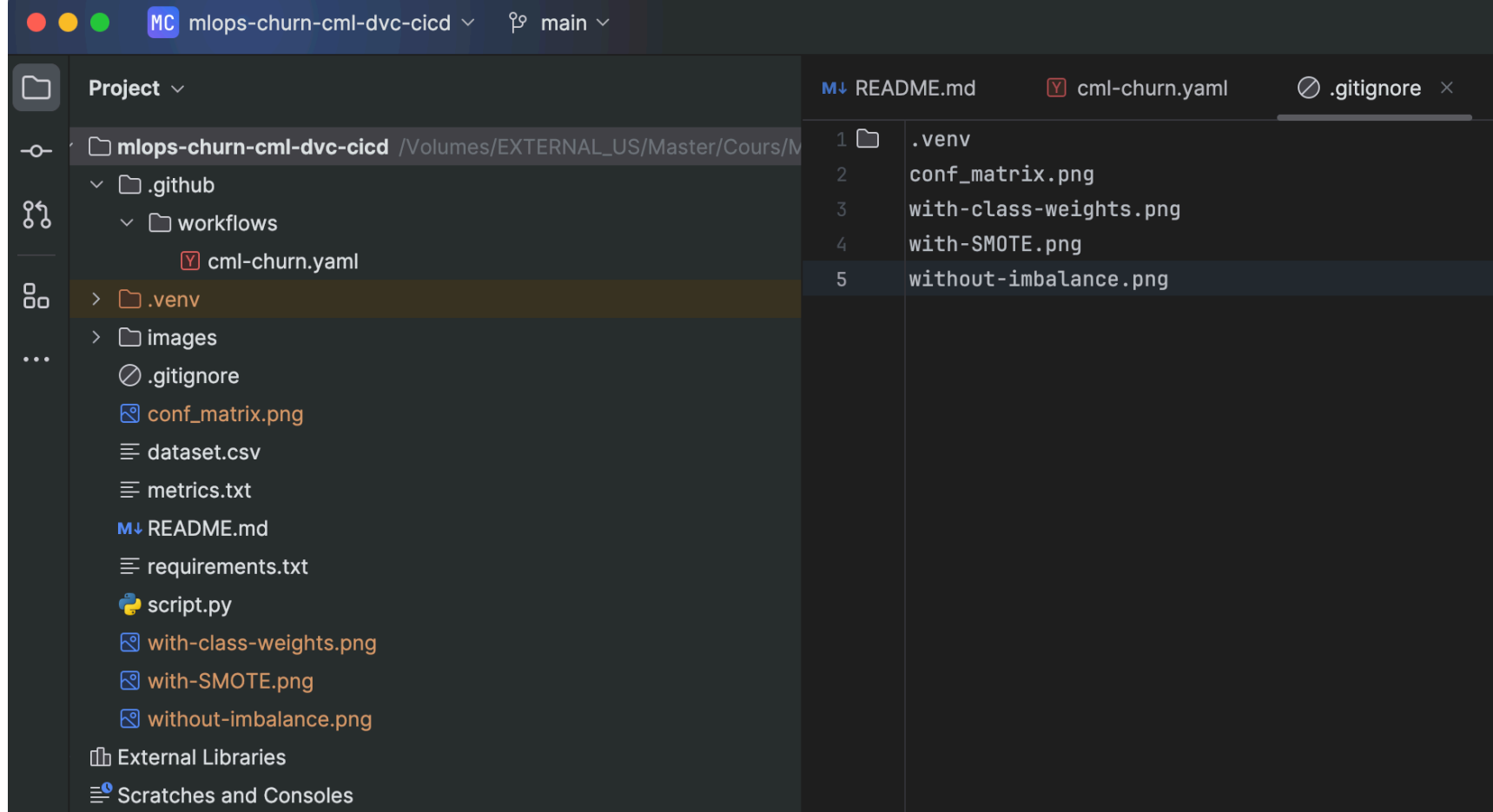
Without Imbalance

Comparaison avec un jeu de données où le déséquilibre de classe n'est pas traité (baseline).



.gitignore

Fichier configuré pour exclure les fichiers temporaires, l'environnement virtuel et les données sensibles du dépôt Git.



GitHub Actions + CML Run

Interface GitHub Actions montrant le succès de l'exécution du workflow CML.

devsahamerlin / mlops-churn-cml-dvc-cicd

Code Issues Pull requests Actions Projects Wiki Security Insights Settings

churn-cml-imbalanced-dataset

first commit #1

Re-run all jobs

Summary

All jobs

run

Run details

Usage

Workflow file

run

succeeded 5 minutes ago in 2m 0s

Search logs




















- Set up job 3s
- Run actions/checkout@v5 0s
- Run actions/setup-node@v6 4s
- Run actions/setup-python@v6 0s
- Run iterative/setup-cml@v2 8s
- Prepare Environment 1m 39s
- Generate Report 3s
- Post Run actions/setup-python@v6 0s
- Post Run actions/setup-node@v6 0s
- Post Run actions/checkout@v5 0s
- Complete job 0s

Run # Get the report

21 <https://github.com/devsahamerlin/mlops-churn-cml-dvc-cicd/commit/4b1904dde888d0b9ead3462d32ec6cd626f6a54a#commitcomment-174990771>

CML Reports <https://github.com/devsahamerlin/mlops-churn-cml-dvc-cicd/commit/4b1904dde888d0b9ead3462d32ec6cd626f6a54a#commitcomment-174990771>

Rapport automatique publié par CML en commentaire du commit, incluant les métriques et la matrice de confusion.

- ✓  .github/workflows
 -  cml-churn.yaml
-  .gitignore
- ✓  .idea
 -  .gitignore
- ✓  inspectionProfiles
 -  Project_Default.xml
 -  profiles_settings.xml
-  misc.xml
-  mlops-churn-cml-dvc-cic...
-  modules.xml
-  vcs.xml
-  README.md
-  dataset.csv
- ✓  images
 -  metrics.png
 -  metrics.txt
 -  requirements.txt
 -  script.py

 **github-actions[bot]** 6 minutes ago

Metrics

RandomForestClassifier without-imbalance F1-score of Training is: 68.60 % F1-Score of Validation is: 56.56 %

RandomForestClassifier with-class-weights F1-score of Training is: 78.97 % F1-Score of Validation is: 63.08 %

RandomForestClassifier with-SMOTE F1-score of Training is: 87.45 % F1-Score of Validation is: 60.82 %

RandomForestClassifier without-imbalance

F1-score of Training is: 68.60 %

F1-Score of Validation is: 56.56 %

RandomForestClassifier with-class-weights

F1-score of Training is: 78.97 %

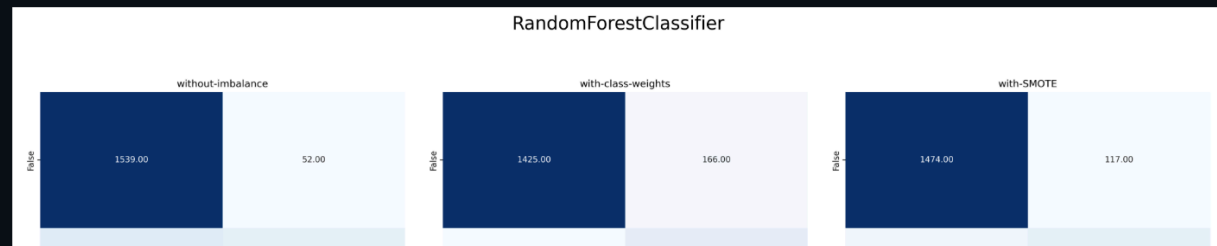
F1-Score of Validation is: 63.08 %

RandomForestClassifier with-SMOTE

F1-score of Training is: 87.45 %

F1-Score of Validation is: 60.82 %

Confusion Matrix



GitHub's collaborations (PR)

Open Pull Request

Création d'une Pull Request pour intégrer les changements de la branche de développement vers la branche principale.

Open

cicd: CML run #1

devsahamerlin wants to merge 1 commit into `main` from `dev`

cicd: CML run

f4923d1

github-actions bot

commented 9 minutes ago

...

Metrics

RandomForestClassifier without-imbalance F1-score of Training is: 68.50 % F1-Score of Validation is: 56.70 %

RandomForestClassifier with-class-weights F1-score of Training is: 78.71 % F1-Score of Validation is: 63.47 %

RandomForestClassifier with-SMOTE F1-score of Training is: 87.95 % F1-Score of Validation is: 60.13 %

RandomForestClassifier without-imbalance

F1-score of Training is: 68.50 %

F1-Score of Validation is: 56.70 %

RandomForestClassifier with-class-weights

F1-score of Training is: 78.71 %

F1-Score of Validation is: 63.47 %

RandomForestClassifier with-SMOTE

F1-score of Training is: 87.95 %

F1-Score of Validation is: 60.13 %

Confusion Matrix

RandomForestClassifier

without-imbalance

with-class-weights

with-SMOTE

No one—[assign yourself](#)

Labels

None yet

Projects

None yet

Milestone

No milestone

Development

Successfully merging this pull request may close these issues.

None yet

Notifications

Customize

Unsubscribe

You're receiving notifications because you authored the thread.

1 participant

Lock conversation

Pull Request List

Vue d'ensemble des Pull Requests ouvertes dans le projet.

The screenshot shows the GitHub interface for the repository `mlops-churn-cml-dvc-cicd`. The **Pull requests** tab is selected, showing 1 open pull request. A notification banner at the top encourages labeling issues and pull requests for new contributors. The pull request list shows one open pull request titled `cicd: CML run` by `devsahamerlin`, opened 11 minutes ago. The interface includes filters, labels, milestones, and a search bar.

Label issues and pull requests for new contributors
Now, GitHub will help potential first-time contributors [discover issues](#) labeled with `good first issue` [Dismiss](#)

Filters Labels 9 Milestones 0 [New pull request](#)

1 Open 0 Closed

	Author	Label	Projects	Milestones	Reviews	Assignee	Sort
<input type="checkbox"/> cicd: CML run ✓							1
#1 opened 11 minutes ago by devsahamerlin							

ProTip! Find everything you created by searching `author:devsahamerlin`.

© 2026 GitHub, Inc. [Terms](#) [Privacy](#) [Security](#) [Status](#) [Community](#) [Docs](#) [Contact](#) [Manage cookies](#) [Do not share my personal information](#)

GitHub Actions (Merged PL)

Exécution automatique du pipeline CI/CD suite à la fusion (merge) d'une Pull Request.

The screenshot shows the GitHub Actions page for the repository `mlops-churn-cml-dvc-cicd`. The **Actions** tab is selected, showing a list of workflow runs. The left sidebar contains links to **Actions**, **churn-cml-imbalanced-dataset**, **Management**, **Caches**, **Attestations**, **Runners**, **Usage metrics**, and **Performance metrics**. The main area displays 3 workflow runs, including a merge pull request, a CML run, and a first commit.

Actions [New workflow](#)

All workflows

churn-cml-imbalanced-dataset

Management

- Caches
- Attestations
- Runners
- Usage metrics
- Performance metrics

All workflows

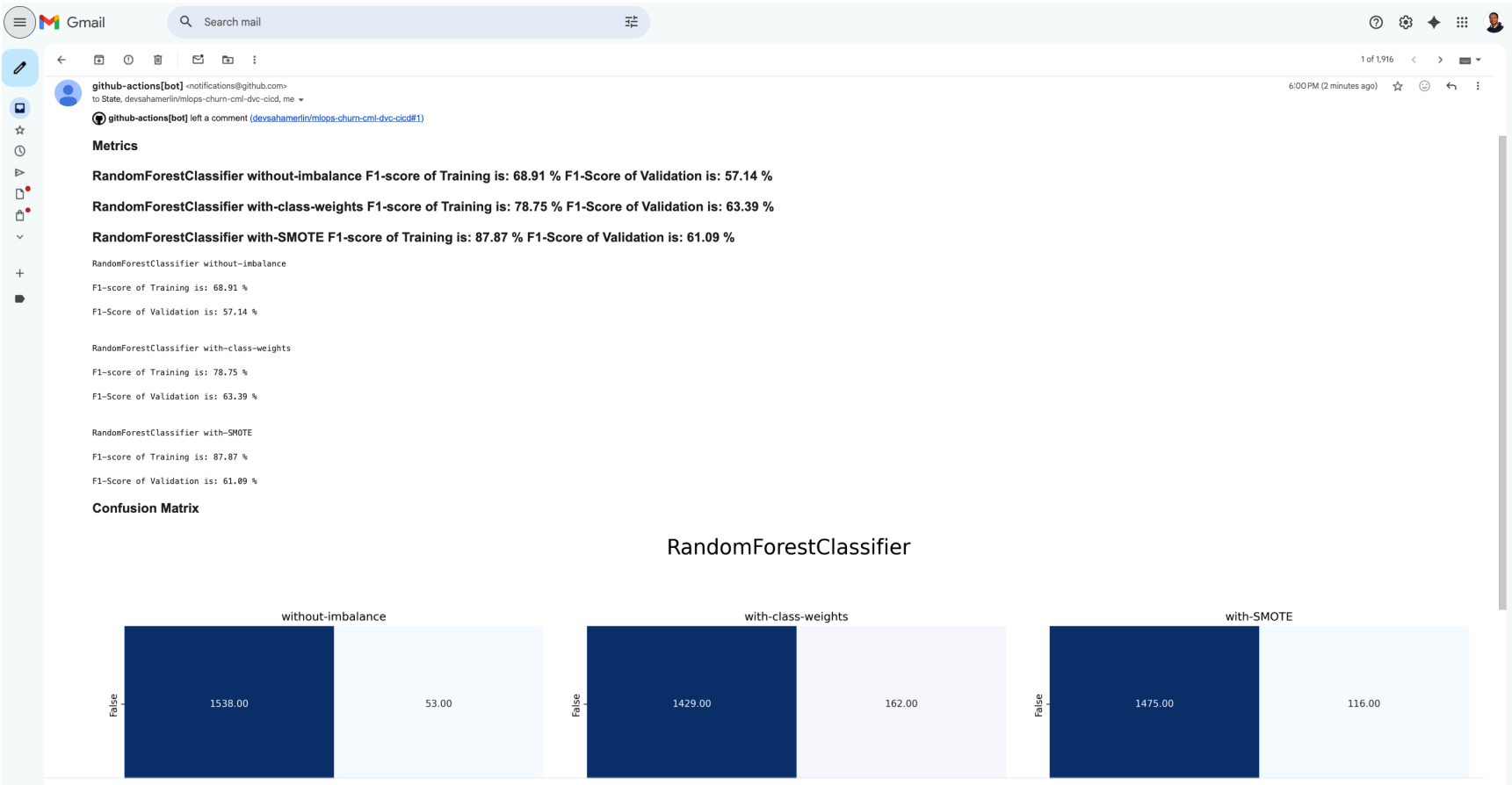
Showing runs from all workflows

3 workflow runs

	Event	Status	Branch	Actor
Merge pull request #1 from devsahamerlin/dev churn-cml-imbalanced-dataset #3: Commit <code>b8660f8</code> pushed by devsahamerlin		<i>In progress</i>	<code>main</code>	now
cicd: CML run churn-cml-imbalanced-dataset #2: Commit <code>f4923d1</code> pushed by devsahamerlin		<i>2m 6s</i>	<code>dev</code>	13 minutes ago
first commit churn-cml-imbalanced-dataset #1: Commit <code>4b1904d</code> pushed by devsahamerlin		<i>2m 4s</i>	<code>main</code>	24 minutes ago

CML E-mail Notifications

Notification par e-mail reçue indiquant le statut de l'exécution du workflow.



CML API rate limit exceeded (5th Run)

API rate limit exceeded

Erreur rencontrée lorsque le nombre de requêtes à l'API GitHub dépasse la limite autorisée (souvent dû à trop d'exécutions rapprochées).

✖ Merge pull request #2 from devsahamerlin/feat/docs #5

Re-run jobs ⌵

⋮

Summary

All jobs

run

Run details

Usage

Workflow file

> Annotations

1 error

run

failed 1 minute ago in 11s

Search logs

Explain error

🔄

⚙️

> ✔ Set up job

2s

> ✔ Run actions/checkout@v5

1s

> ✔ Run actions/setup-node@v6

3s

> ✔ Run actions/setup-python@v6

0s

⌵ ✖ Run iterative/setup-cml@v2

1s

1 ▶ Run iterative/setup-cml@v2

12 **Error:** HttpError: API rate limit exceeded for 25. (But here's the good news: Authenticated requests get a higher rate limit. Check out the documentation for more details.)

🕒 Prepare Environment

0s

🕒 Generate Report

0s

🕒 Post Run actions/setup-python@v6

0s

🕒 Post Run actions/setup-node@v6

0s

> ✔ Post Run actions/checkout@v5

0s

> ✔ Complete job

0s

Reduce Pipeline Run

- Reduce GitHub Actions Pipeline Run: Run only on PR to Main and PR to DEV

```
on:
  push:
    branches: [ main ]
  pull_request:
    branches: [ main, dev ]
```

Git log

Visualisation de l'historique Git montrant les branches, les commits et les fusions.

```
(.venv) merlinsaha@Mac mlops-churn-cml-dvc-cicd % git log --oneline --graph --decorate --all
* 97a8de8 (docs/goal) docs: improve documentation
| * 3c64b8a (HEAD -> main, origin/main, origin/HEAD) Merge pull request #5 from devsaahamerlin/dev
| | \
| | /
| / |
| / |
* | f60187a (origin/dev, dev) Merge pull request #4 from devsaahamerlin/docs/add-git-graph
| \ \
| * | 7a71b88 (origin/docs/add-git-graph, docs/add-git-graph) doc: add git graph
| / /
* | 0033a6b Merge pull request #3 from devsaahamerlin/refactor/cml-cicd
| \ \
| * | 02fa728 (origin/refactor/cml-cicd, refactor/cml-cicd) refactor: run pipeline only on main and dev
| / /
* | aa88265 Merge pull request #2 from devsaahamerlin/feat/docs
| \ \
| * | 1c664d7 (origin/feat/docs, feat/docs) docs: update README documentations
| / /
| * b8660f8 Merge pull request #1 from devsaahamerlin/dev
| | \
| | /
| / |
| / |
* | f4923d1 cicd: CML run
| /
* 4b1904d first commit
(.venv) merlinsaha@Mac mlops-churn-cml-dvc-cicd %
```

Résumé

Nous avons :

- automatisé l'entraînement du modèle à chaque push sur la main et chaque PR sur la main et la dev,
- généré automatiquement les métriques (metrics.txt),
- généré automatiquement la matrice de confusion (conf_matrix.png),
- publié un rapport lisible directement dans GitHub (CML).

C'est déjà une CI ML : mesure continue, feedback centralisé, traçabilité.

Limites actuelles

- Données dans le repo (dataset.csv) : pas scalable, pas sécurisé.
- Modèle entraîné non archivé automatiquement sur stockage externe.
- Pas de versioning propre pour les artefacts lourds (.pkl, gros CSV, etc.).
- Runner sur un CPU et non GPU, bien pour les petits model et dataset

Conclusion

Nous venons de construire une boucle *entraînement + reporting* entièrement automatisée via GitHub Actions et CML : *Mesurer en continu*. Dans l'atelier suivant, on ajoute *stocker et versionner en continu*.