UA3 – Travail pratique / Devoir #2

Déploiement cloud de l’API de prédiction du diabète

Nom : Innocent Niyobuhungiro

Cours : UA3 – Méthodologie IA

Date : 2025-12-03

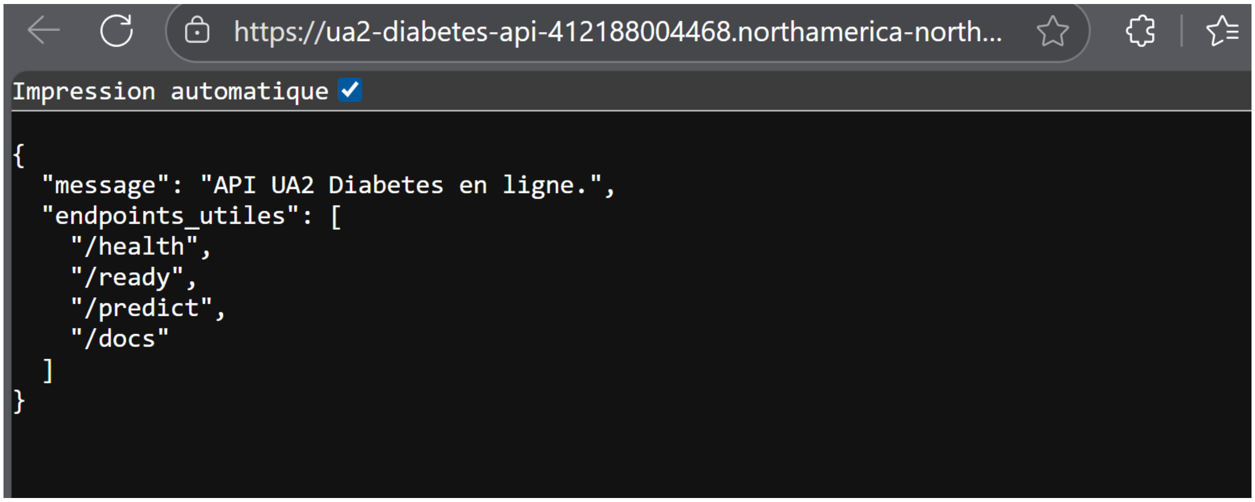
1. URL publique de l’API

L’API de prédiction du diabète déployée sur Google Cloud Run est accessible à l’adresse suivante :

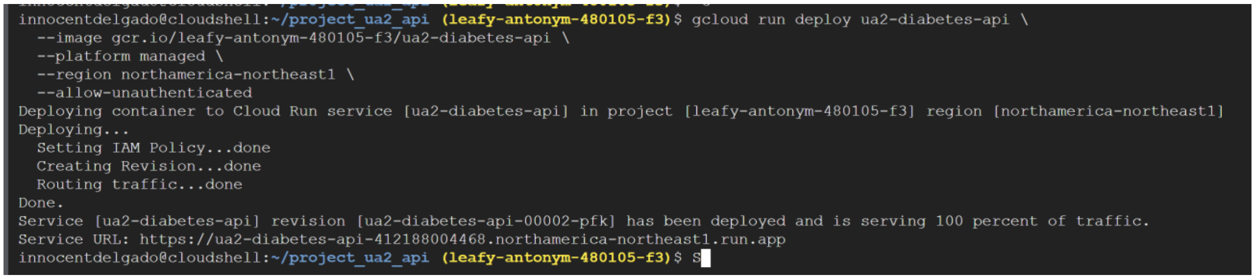
https://ua2-diabetes-api-412188004468.northamerica-northeast1.run.app

L’endpoint principal de prédiction est : /predict

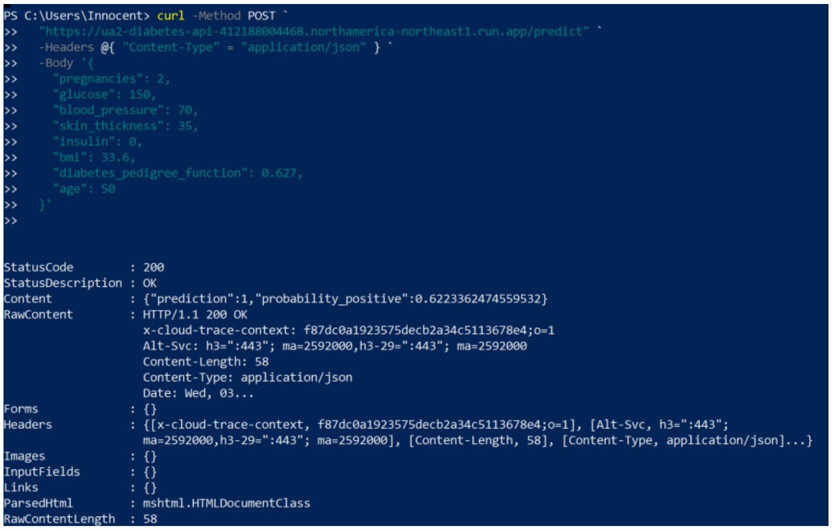
2. Capture d’écran de la requête de test



```
{
  "message": "API UA2 Diabetes en ligne.",
  "endpoints_utiles": [
    "/health",
    "/ready",
    "/predict",
    "/docs"
  ]
}
```

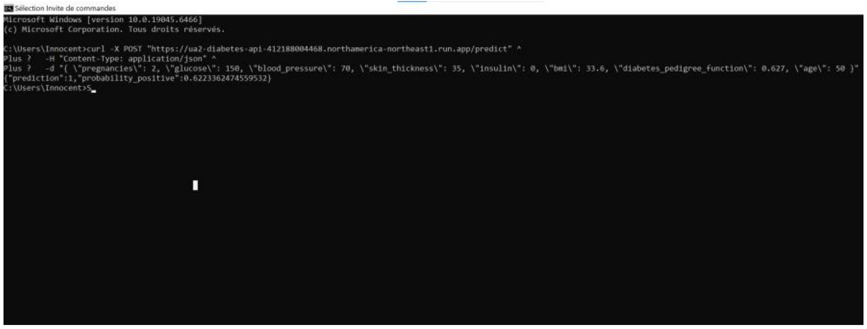


```
innocentdelgado@cloudshell:~/project_ua2_api (leafy-antonym-480105-f3)$ gcloud run deploy ua2-diabetes-api \
  --image gcr.io/leafy-antonym-480105-f3/ua2-diabetes-api \
  --platform managed \
  --region northamerica-northeast1 \
  --allow-unauthenticated
Deploying container to Cloud Run service [ua2-diabetes-api] in project [leafy-antonym-480105-f3] region [northamerica-northeast1]
Deploying...
  Setting IAM Policy...done
  Creating Revision...done
  Routing traffic...done
Done.
Service [ua2-diabetes-api] revision [ua2-diabetes-api-00002-pfk] has been deployed and is serving 100 percent of traffic.
Service URL: https://ua2-diabetes-api-412188004468.northamerica-northeast1.run.app
innocentdelgado@cloudshell:~/project_ua2_api (leafy-antonym-480105-f3)$ S
```



```
PS C:\Users\Innocent> curl -Method POST `
>>   "https://ua2-diabetes-api-412188004468.northamerica-northeast1.run.app/predict" `
>>   -Headers @{ "Content-Type" = "application/json" } `
>>   -Body '{
>>     "pregnancies": 2,
>>     "glucose": 150,
>>     "blood_pressure": 70,
>>     "skin_thickness": 35,
>>     "insulin": 0,
>>     "bmi": 33.6,
>>     "diabetes_pedigree_function": 0.627,
>>     "age": 50
>>   }'
>>


StatusCode        : 200
StatusDescription : OK
Content           : {"prediction":1,"probability_positive":0.6223362474559532}
RawContent        : HTTP/1.1 200 OK
                    x-cloud-trace-context: f87dc0a1923575decb2a34c5113678e4;o=1
                    Alt-Svc: h3=":443"; ma=2592000,h3-29=":443"; ma=2592000
                    Content-Length: 58
                    Content-Type: application/json
                    Date: Wed, 03...
Forms             : {}
Headers           : {[x-cloud-trace-context, f87dc0a1923575decb2a34c5113678e4;o=1], [Alt-Svc, h3=":443";
                    ma=2592000,h3-29=":443"; ma=2592000], [Content-Length, 58], [Content-Type, application/json]...}
Images            : {}
InputFields       : {}
Links             : {}
ParsedHtml        : mshtml.HTMLDocumentClass
RawContentLength  : 58
```



```
Microsoft Windows [version 10.0.19045.6466]
(c) Microsoft Corporation. Tous droits réservés.

C:\Users\Innocent>curl -X POST "https://ua2-diabetes-api-412188004468.northamerica-northeast1.run.app/predict" ^
Plus ?   -H "Content-Type: application/json" ^
Plus ?   -d "{ \"pregnancies\": 2, \"glucose\": 150, \"blood_pressure\": 70, \"skin_thickness\": 35, \"insulin\": 0, \"bmi\": 33.6, \"diabetes_pedigree_function\": 0.627, \"age\": 50 }"
{"prediction":1,"probability_positive":0.6223362474559532}
C:\Users\Innocent>S
```

3. Brève description de l’architecture (Docker + Cloud Run)

L’API est développée avec FastAPI et charge un modèle de Gradient Boosting sauvegardé dans le fichier best_model_GB_opt.joblib pour prédire le risque de diabète à partir de variables cliniques. L’application est empaquetée dans une image Docker construite avec Google Cloud Build, puis stockée dans Container Registry sous le nom gcr.io/leafy-antonym-480105-f3/ua2-diabetes-api. Cette image est ensuite déployée sur Google Cloud Run en mode fully managed, ce qui permet d’exposer publiquement les endpoints /health, /ready, /predict et /docs via l’URL de service ci-dessus. L’accès est configuré en "allow-unauthenticated", ce qui permet d’appeler l’API directement depuis n’importe quel client HTTP (curl, Postman, navigateur).