

# 7. PROJECT KICK OFF

LEV KIWI

# PROJET EN DATA SCIENCE

## Introduction

Prenez connaissance du dataset

<https://www.kaggle.com/competitions/house-prices-advanced-regression-techniques/overview>

# House Prices - Advanced Regression Techniques

Predict sales prices and practice feature engineering, RFs, and gradient boosting

Overview Data Code Models Discussion Leaderboard Rules

## Overview

∞ This competition runs indefinitely with a rolling leaderboard. [Learn more.](#)

## Description

### Start here if...

You have some experience with R or Python and machine learning basics. This is a perfect competition for data science students who have completed an online course in machine learning and are looking to expand their skill set before trying a featured competition.

### Getting Started Notebook

To get started quickly, feel free to take advantage of [this starter notebook](#).

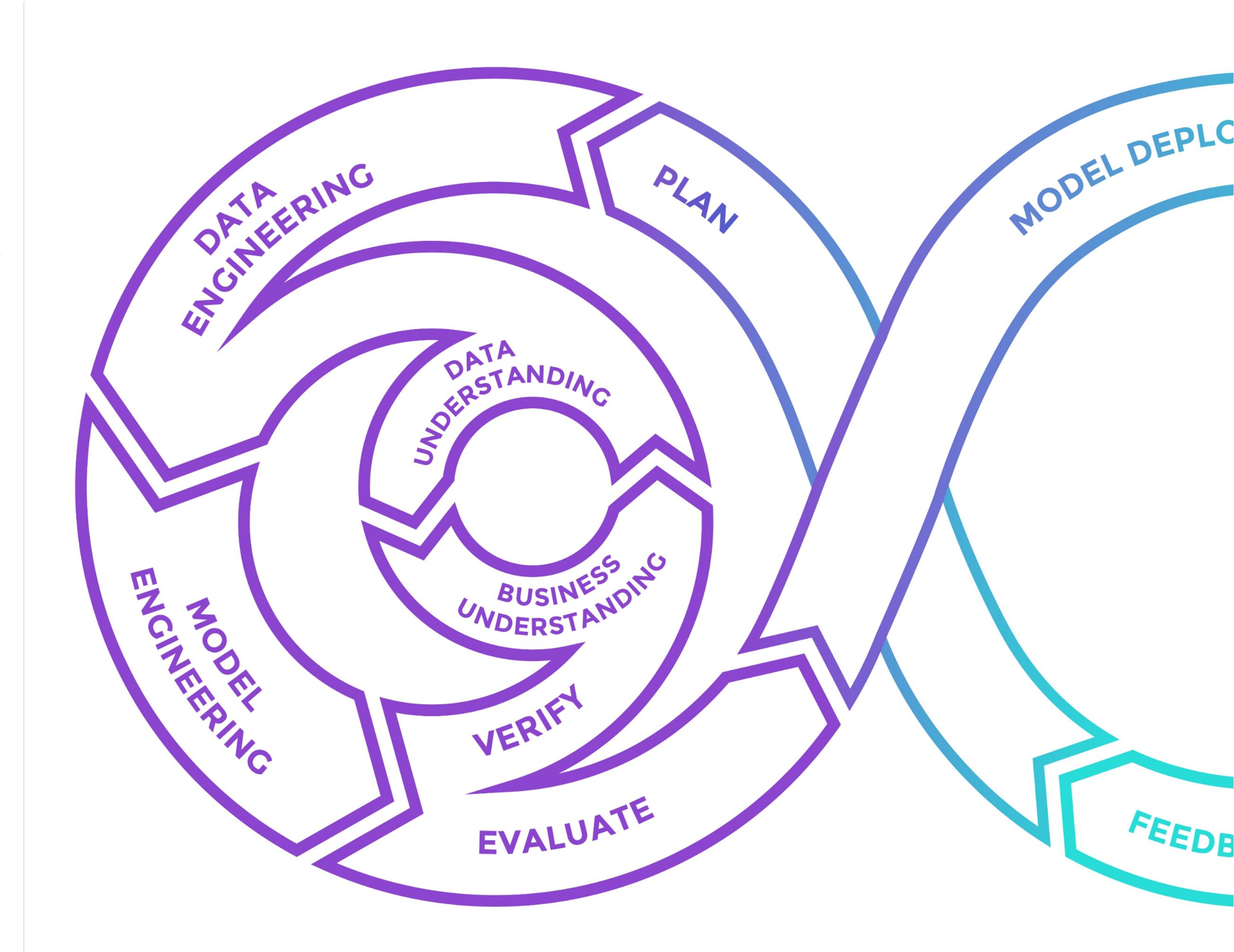
## Competition Description



# CONSIGNES POUR LE PROJET

## Vos tâches

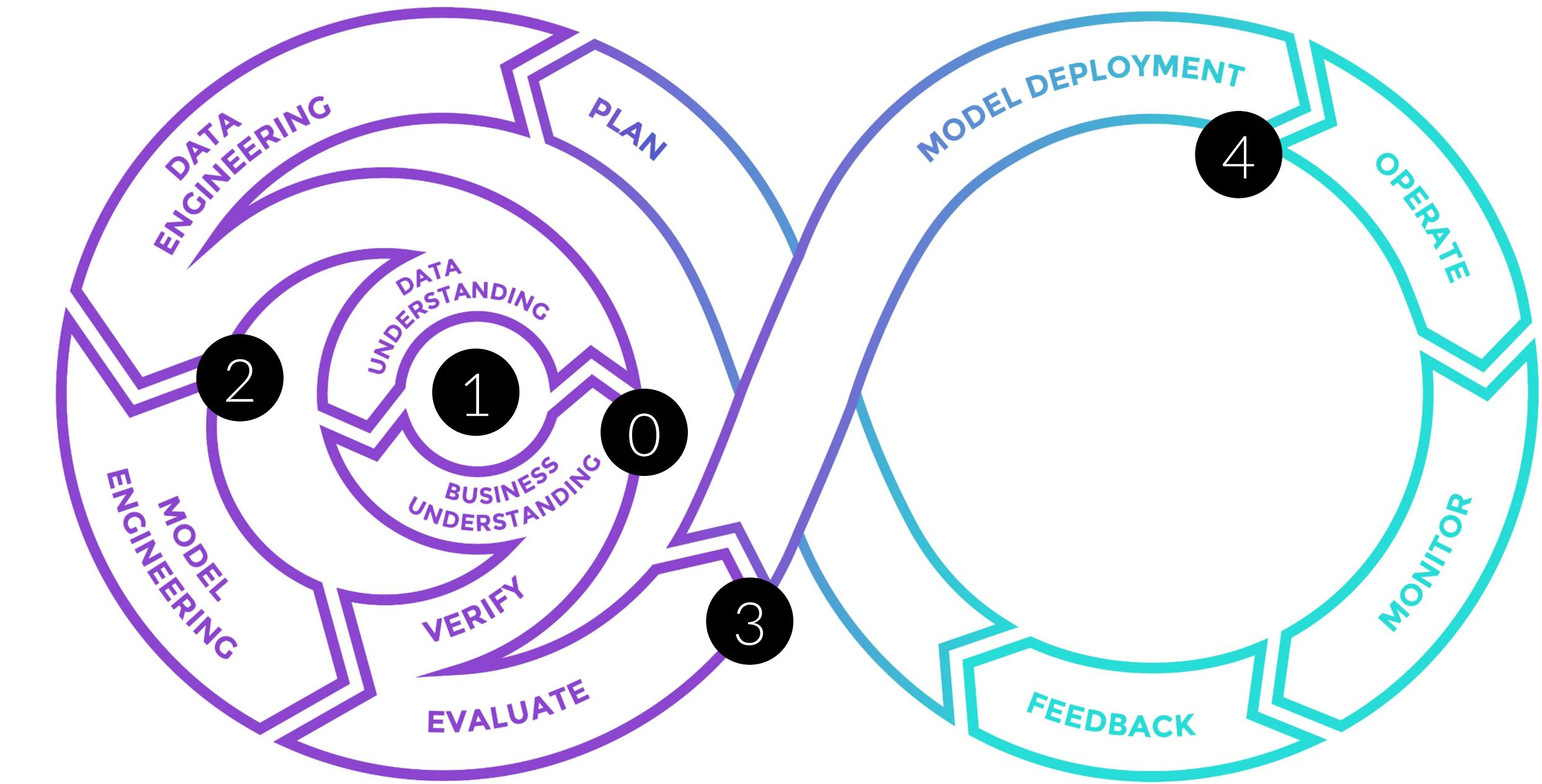
- Pour ce projet, vous incarnez une équipe de data scientists.
- Vous produirez un **use case réaliste** avec le dataset à disposition.
- Vous irez à travers toutes les étapes du **CRISP-ML**, i.e. de la phase d'idéation jusqu'au déploiement (théorique).
- Le déploiement correspond à la **soumission de votre modèle** ML sur [kaggle.com](https://www.kaggle.com)
- Vous rendrez un **Jupyter Notebook** contenant votre code ainsi qu'une documentation exhaustive expliquant comment vous êtes arrivé à votre modèle.



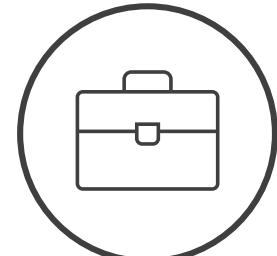
# TIMELINE SUGGÉRÉE

## Phases de projet

Planning conseillé sur la fin du semestre.

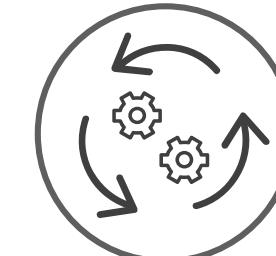


0. Phase zéro



Semaines 7-8

1. Phase d'idéation



Semaines 9 - 10

2. Phase de design



Semaines 11-14

3. Blueprint



Semaine 15

4. Rendu



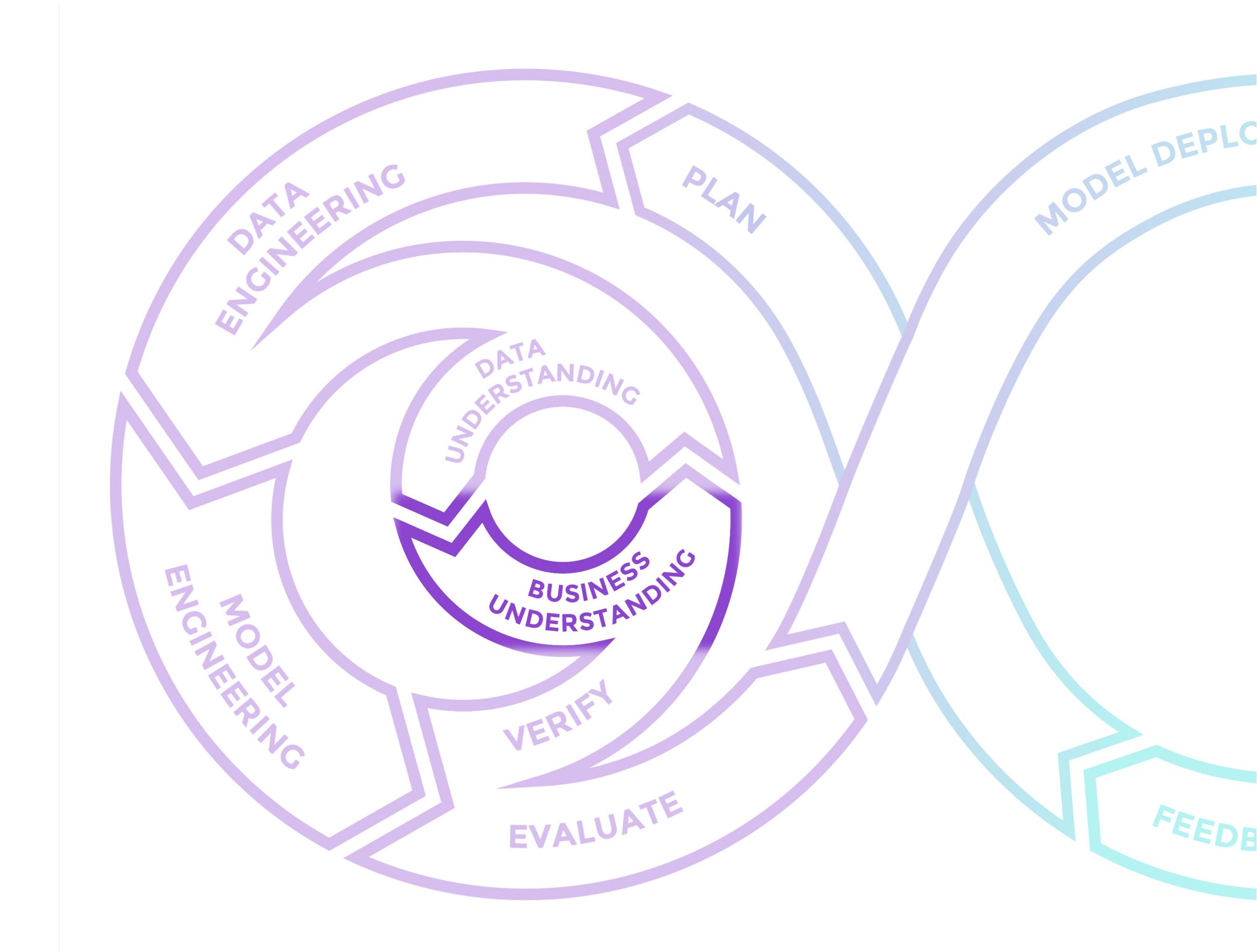
Fin des cours

# PHASE ZÉRO

## Produire un use case

Commencez par former les équipes et produire un use case réaliste avec les dataset à disposition. Remplissez le ML Canvas de manière aussi détaillée que possible.

<https://ml-ops.org/content/phase-zero>

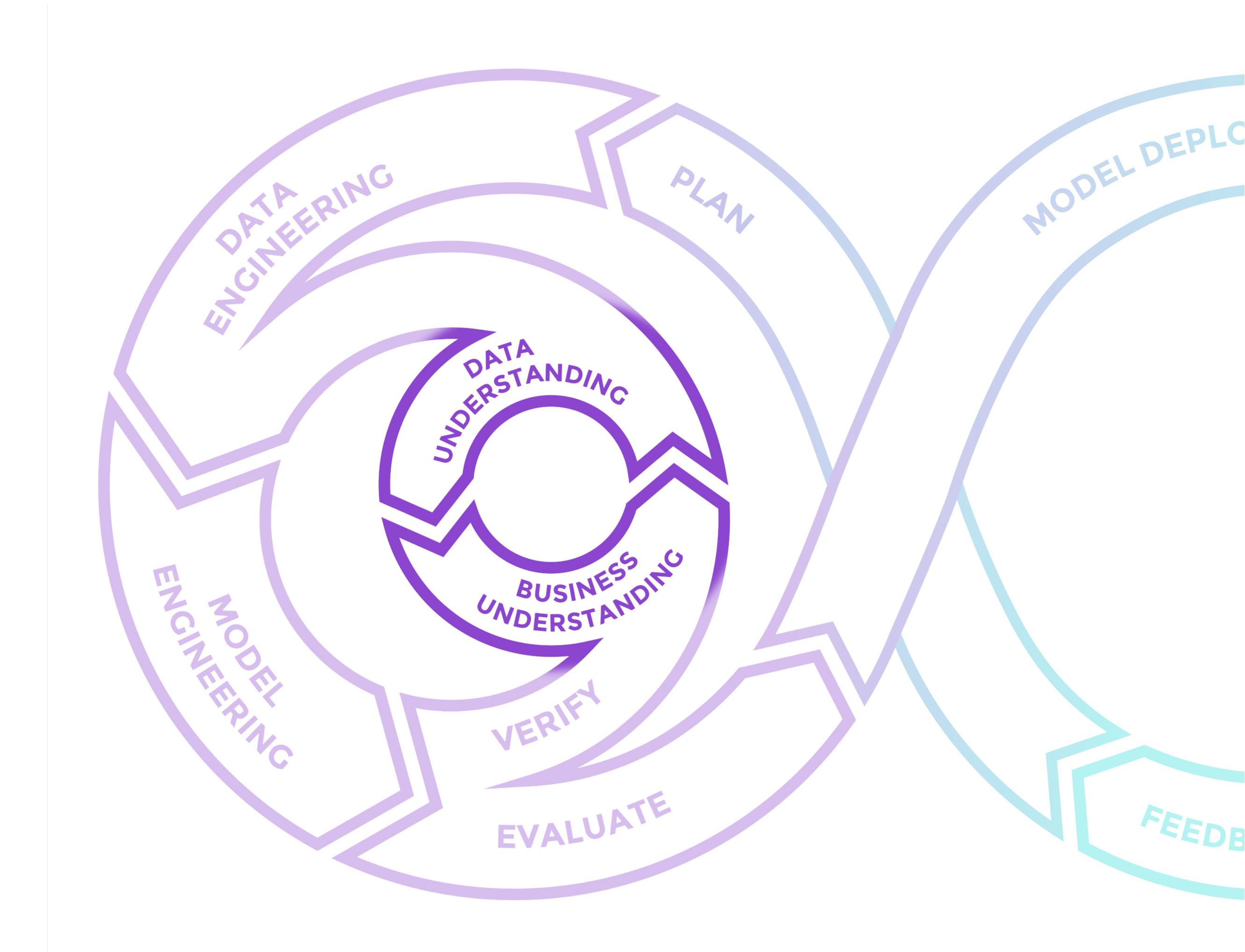


# PHASE D'IDÉATION

## Data Anaysis & Data Mining

Analysez les données en votre possession. Emettez des **hypothèses** sur les données qui possèdent le plus d'information selon vous. Créez un **modèle baseline** avec quelques variables sélectionnées par vos soins. Justifiez le raisonnement qui vous a conduit à ce modèle.

<https://ml-ops.org/content/crisp-ml>



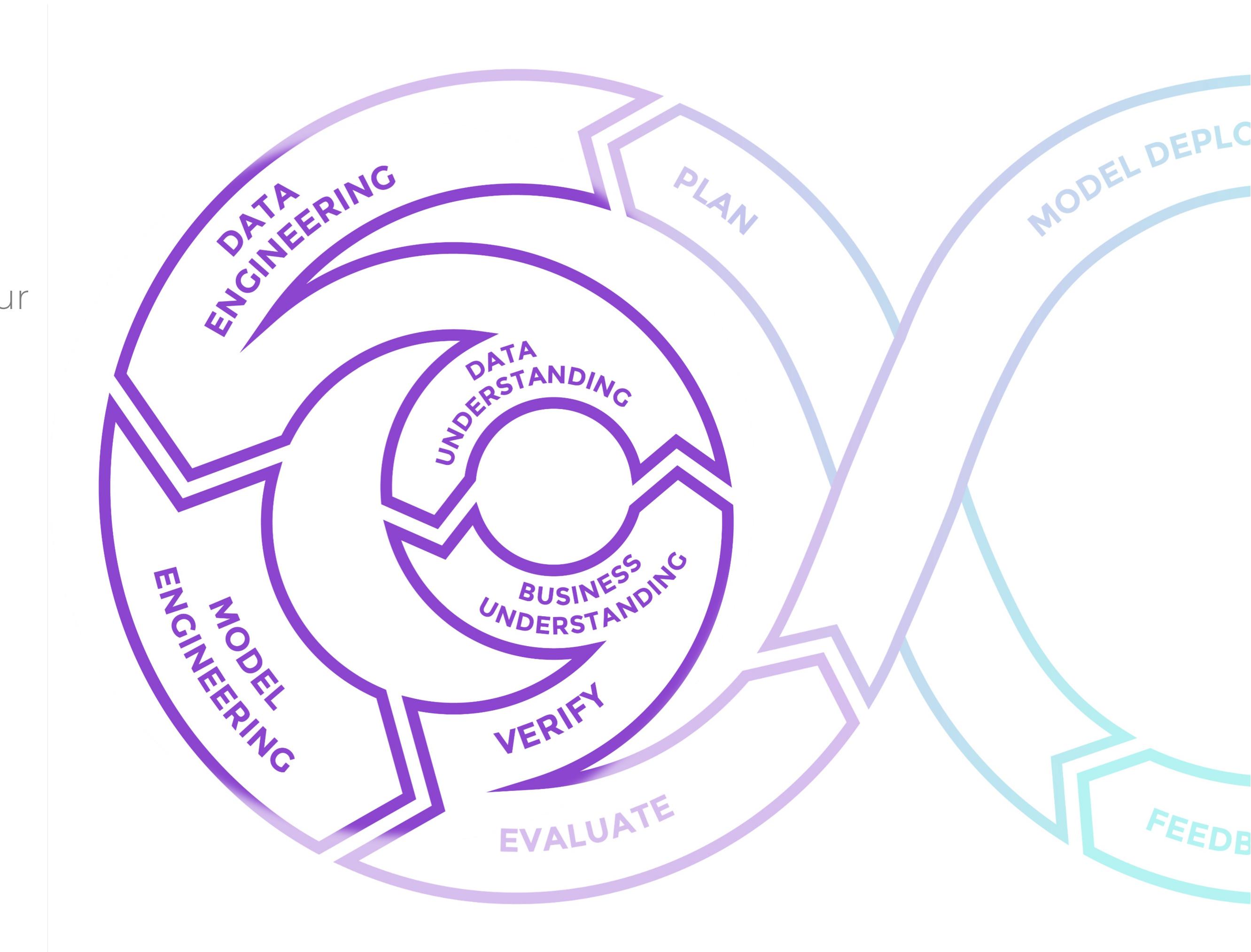
# DESIGN PHASE

## Data Engineering

Préparez une **Scikit-learn pipeline** afin de traiter les données du dataset train et test. La pipeline doit être capable de **fit()** sur le dataset **train.csv** et **transform()** sur le dataset **test.csv**.

## Model Engineering

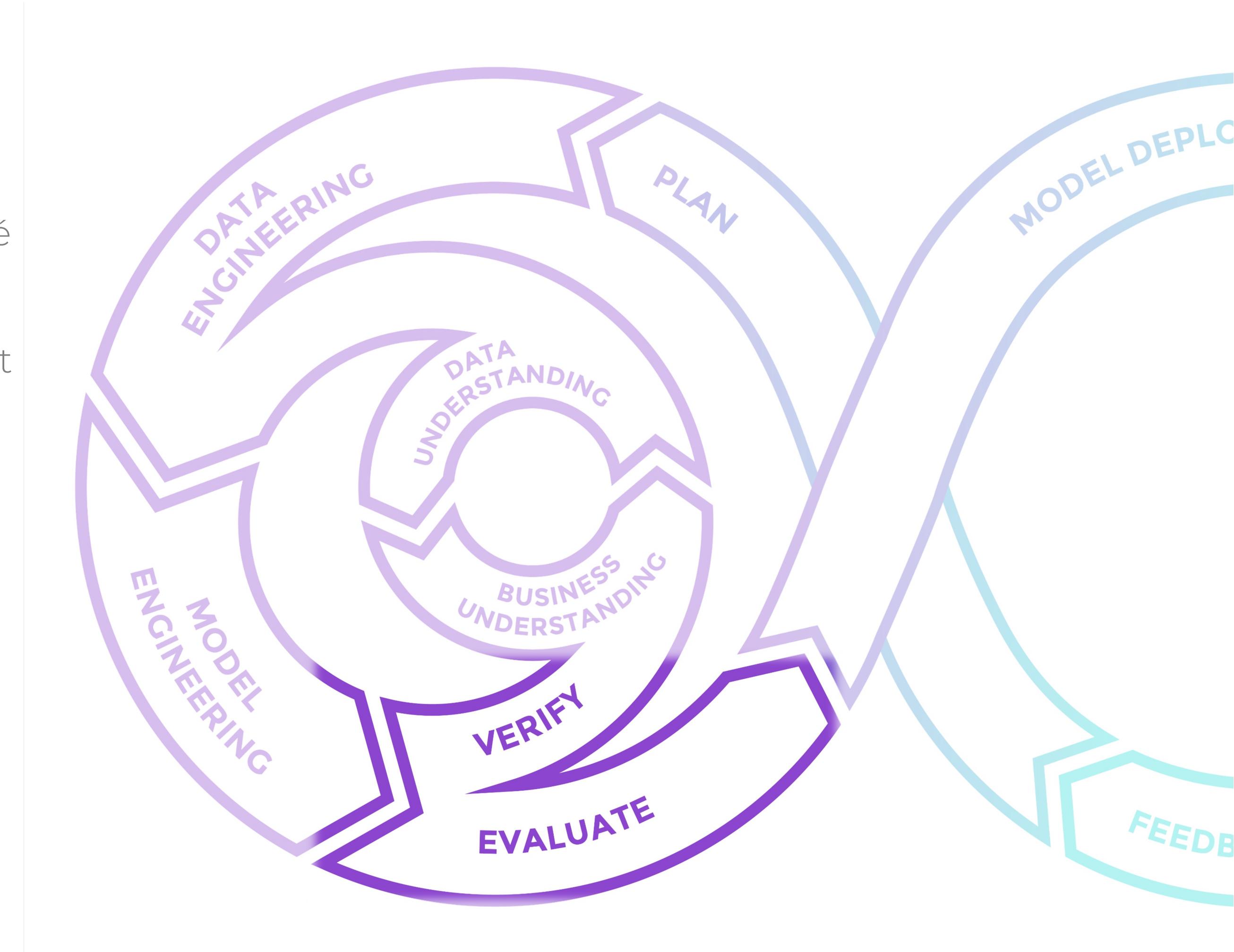
Testez des modèles pertinents pour votre use case. Les **hyperparamètres** doivent être ajustés. La performance des modèles est estimée indépendamment de la soumission sur kaggle. Le modèle est **interprété** à la lumière de votre compréhension du métier.



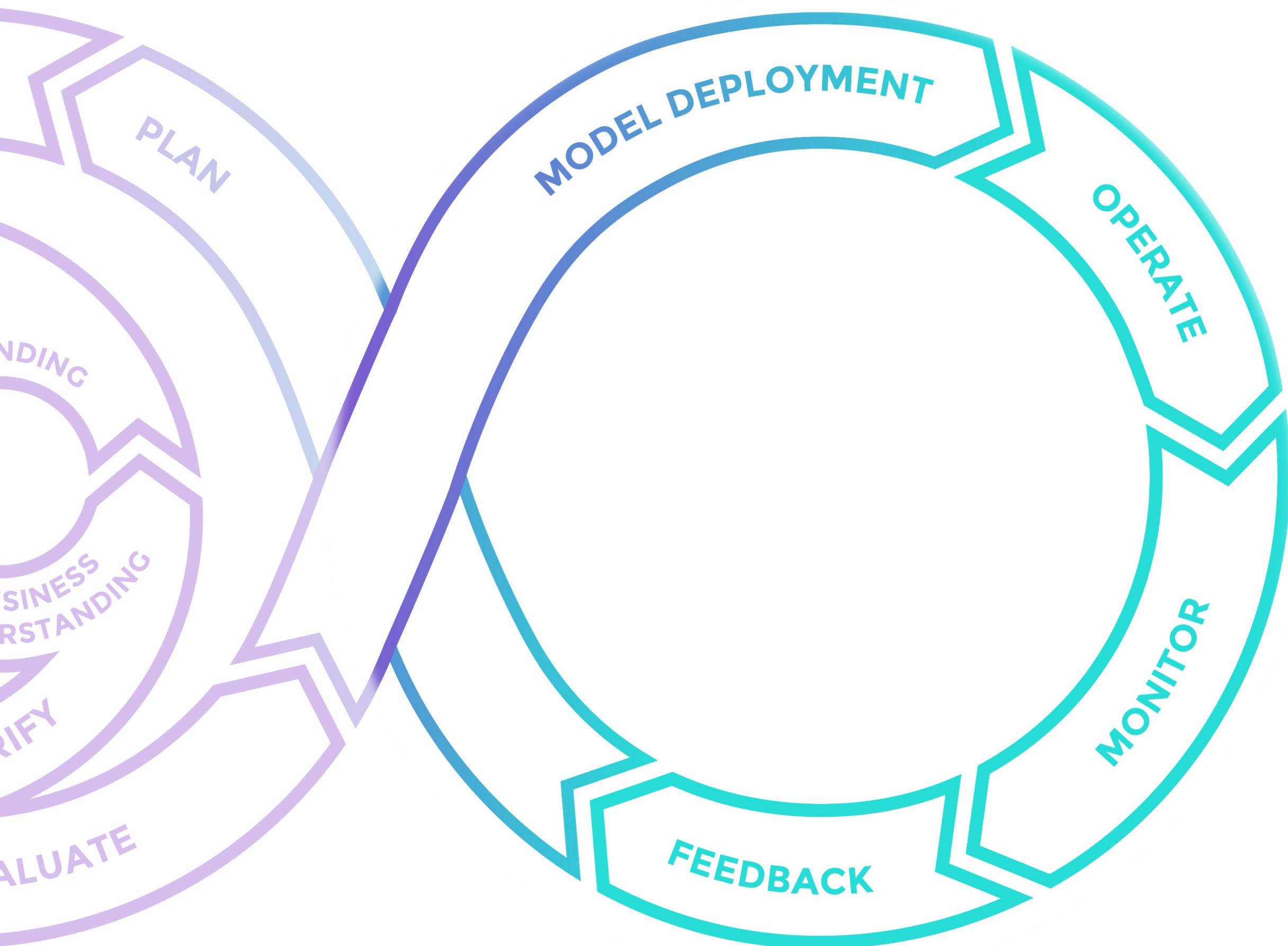
# DESIGN PHASE - ENDING

## Evaluate

Sélectionnez le modèle que vous estimez être le plus approprié pour votre use case. Discutez les raisons pour lesquelles vous avez sélectionné ce modèle et **interprétez** son fonctionnement sur votre dataset. Discutez **l'efficacité** et les **limites** du modèle pour votre business case. Préparez la **pipeline** complète à déployer (preprocessing & modèle), **fit()** sur le **train.csv** et **predict()** sur le **test.csv**.



# BLUEPRINT & OPERATION PHASE



## ML en production

Soumettez vos résultats sur kaggle, reportez vos résultats dans le Jupyter Notebook et **documentez** votre code afin de raconter une histoire. Vous devez convaincre le management que votre solution va fonctionner en production. Discutez comment vous allez **maintenir** et **monitorez** l'algorithme ainsi que recueillir le feedback de vos utilisateurs. Finalement, discutez des potentiels **implications éthiques et légales** de votre solution.

<https://ml-ops.org/content/mlops-principles>



**SOUMETTRE SUR CYBERLEARN VOTRE JUPYTER  
NOTEBOOK COMPILEÉ AVEC LE FILENAME**

**GroupCode-Nom1\_Nom2\_Nom3.ipynb**