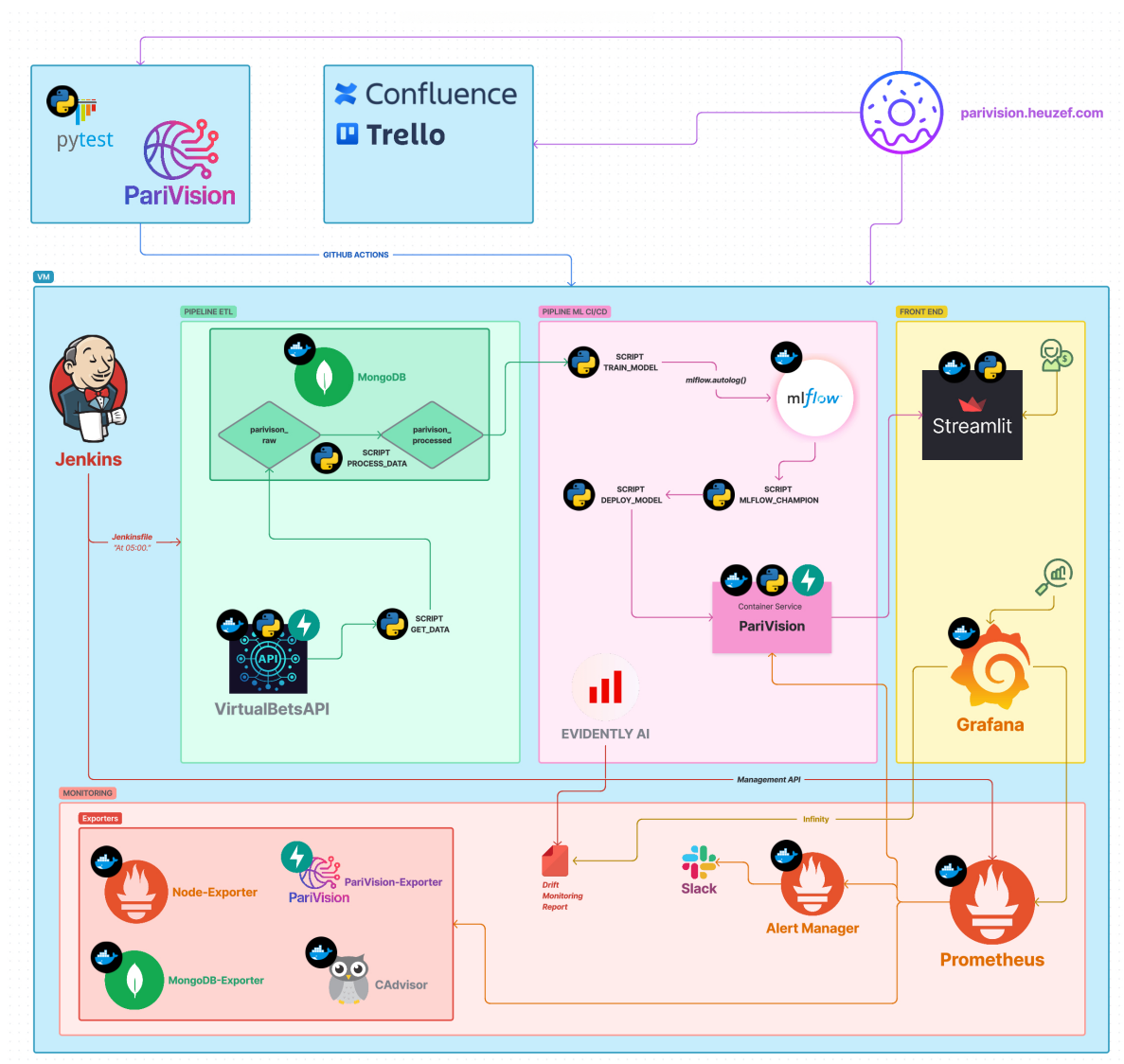


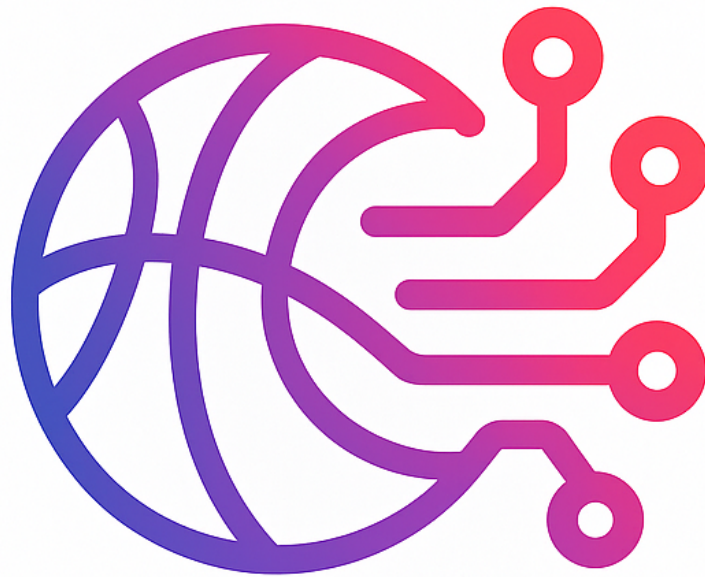


1. Introduction

- **Projet PariVision** : prédiction des résultats des matchs de Basketball (NBA)
- Introduction
- Objectifs
- Contribution à l'industrie
- Principaux défis identifiés
- Flux de travail
- Hébergement du projet

Projet PariVision : prédiction des résultats des matchs de Basketball (NBA)





PariVision

Introduction

Le basketball, en particulier la NBA (National Basketball Association), figure parmi les sports les plus prisés au monde. Il attire des millions de spectateurs et génère un marché économique colossal, notamment dans le domaine des paris sportifs. Avec une saison régulière comprenant 82 matchs par équipe et des playoffs intenses, la NBA offre un volume conséquent de données exploitables pour l'analyse prédictive. Dans ce contexte, où chaque tir peut influencer l'issue d'un match et générer des enjeux stratégiques, économiques (paris sportifs, droits TV) ou émotionnels (engagement des fans), la capacité à prédire la réussite d'un tir spécifique devient un atout précieux.

Notre projet **PariVision** se concentre sur cette problématique en combinant machine learning et ingénierie MLOps pour concevoir un système prédictif robuste et automatisé, capable d'estimer les

résultats des matchs NBA. La complexité réside dans la modélisation des multiples facteurs influençant un tir : position sur le terrain, pression défensive, ou même le contexte du match (dernières secondes, enjeux du score). Les approches statiques peinent à s'adapter à la variabilité des données sportives, nécessitant une infrastructure capable d'évoluer avec les saisons et les métriques émergentes.

La prédiction des tirs réussis au basketball est un enjeu central pour l'analyse de la performance des joueurs, l'optimisation des stratégies d'équipe et l'engagement des fans. Dans ce contexte, le projet « Prédiction des tirs réussis des joueurs NBA » vise à concevoir, entraîner et déployer un modèle de machine learning capable de prédire, à partir des caractéristiques du tir et de son contexte, la probabilité de réussite d'un tir tenté lors d'un match NBA.

Ce projet s'inscrit donc dans une démarche complète de science des données, de la préparation des données brutes jusqu'au déploiement et à la surveillance du modèle. Il offre un cadre réutilisable et évolutif pour l'analyse prédictive dans le domaine sportif et au-delà et permet d'illustrer comment le MLOps peut transformer des données brutes en entrées exploitable, en alignant précision algorithmique et exigences opérationnelles.

En résumé, ce projet répond à des besoins concrets de l'industrie du sport, s'appuie sur les technologies les plus récentes en data science et MLOps et anticipe les défis techniques, organisationnels et éthiques liées à l'analyse avancée de la performance sportive.

Objectifs

En ciblant cette problématique, nous élaborons les objectifs suivants :

- **Prédiction contextuelle** : déterminer la probabilité de réussite d'un tir en temps réel, en s'appuyant sur des données synthétiques ou historiques.
- **Automatisation du flux de travail** : conteneuriser chaque étape et orchestrer les pipelines.
- **Adaptabilité continue** : implémenter un système de ré-entraînement conditionnel, où un nouveau modèle ne remplace l'ancien que si son exactitude dépasse celle de l'ancien modèle.
- **Garantir la traçabilité et la reproductibilité** : des intégrations CI/CD, un suivi des performances, la reproductibilité et la sélection optimale des modèles.
- **Surveillance en production** : **assurer** une supervision granulaire des requêtes API, de la base de données ainsi qu'une toute dérive des données ou baisse de performance en production pour ainsi maintenir la fiabilité du modèle dans le temps.

Les principaux résultats attendus sont :

- **Un modèle de machine learning** validé par des métriques robustes.
- **Un service de prédiction déployé** sous forme d'API REST.
- **Une application** d'aide à la décision.

- **Des rapports et panneaux de supervision** détectant la dérive des données ou de la performance, garantissant la fiabilité continue du modèle.
- **Une documentation complète** du pipeline, des choix de modélisation et des résultats, facilitant la réutilisation et l'évolution du projet.

Contribution à l'industrie

- **Amélioration de la performance sportive** : Les analyses issues du modèle permettent d'identifier les zones et situations à haut potentiel de réussite, d'ajuster les stratégies de tir et d'individualiser l'entraînement des joueurs.
- **Appui à la prise de décision en temps réel** : Les équipes NBA utilisent de plus en plus ces outils pour adapter les tactiques pendant les matchs, optimiser les rotations ou anticiper les faiblesses adverses.
- **Valorisation de la data dans l'écosystème sportif** : Le projet illustre comment la collecte, l'analyse et l'exploitation intelligente des données deviennent des leviers majeurs de compétitivité et d'innovation dans le sport professionnel.
- **Transférabilité** : Ce type de pipeline peut être adapté à d'autres sports ou contextes (recrutement, prévention des blessures, expérience fan, etc, ...).

Principaux défis identifiés

- **Qualité et hétérogénéité des données** : Les données sportives sont souvent bruitées, incomplètes ou issues de sources multiples, ce qui nécessite un pré-traitement rigoureux.
- **Évolution des stratégies et du jeu** : Les modèles doivent s'adapter à des changements de règles, de tactiques ou de comportements des joueurs, rendant la supervision essentielle.
- **Dérive des données et des performances** : Sans surveillance, un modèle performant peut devenir obsolète si la distribution des données évolue (ex : nouveaux schémas de jeu, évolution du profil des joueurs).
- **Défis éthiques et réglementaires** : Respect de la vie privée des joueurs, sécurité des données, équité dans l'accès aux technologies avancées.
- **Intégration et adoption** : Faire accepter et intégrer ces outils dans les routines des équipes et du staff technique peut nécessiter un accompagnement au changement.

Flux de travail

Dans une logique MLOps moderne, le projet intègre une pipeline ML Ci/CD :

- **Le suivi d'expériences avec MLflow**, pour garantir la traçabilité, la reproductibilité et la sélection optimale des modèles.
- **Le déploiement du modèle avec FastAPI**, permettant de rendre la prédiction accessible via une API REST.
- **La surveillance continu avec Evidently**, pour détecter toute dérive des données et ainsi maintenir la fiabilité du modèle sur la durée.

Hébergement du projet

Le projet est entièrement hébergé sur un serveur unique en mode monolithique. Bien que cette approche ne soit pas la plus optimale en termes de performance et de scalabilité, elle a été choisie en raison de contraintes budgétaires et de ressources disponibles.

Cette configuration permet de centraliser les opérations et de simplifier la gestion des infrastructures, tout en respectant les limitations financières et techniques.

- Accès a la page d'accueil du projet : [PariVision 2025](#)
- Accès au dépôt Git du projet : https://github.com/DataScientest-Studio/DEC24_MLOPS_PARIS_SPORTIFS/
- Miroir : https://github.com/heuzef/DEC24_MLOPS_PARIS_SPORTIFS

📅 Gestion



Documentation (Confluence)



Kanban (Trello)



Dépôt GIT (Github)

📦 Orchestration



Portainer (9443)



Jenkins (8080)

☁ Services



PariVision API (3000)



VirtualBetsApi (8800)



MLFlow (5000)



Streamlit (8501)

👁 Surveillance



Prometheus (9090)



AlertManager (9093)



Node-Exporter (9100)



MongoDB-Exporter (9216)



PariVision-Exporter (3000)



Grafana (3333)



CAvior (8082)



Drift Monitoring Reports (8000)