



MoneyBall Reloaded



Aksel Volle, Hugo Degeorges, Julien
Nicolas, Corentin Bel, Elie Lévy



Problématique Métier

- Faciliter le recrutement et le remplacement en trouvant des joueurs similaires
 - Pouvoir comparer les joueurs
 - Essayer de catégoriser les joueurs
-

Technologies utilisées

- Python
- Scikit-Learn
- Jupyter Notebook
- Matplotlib
- Streamlit



Quelles données?

- **Saisons 2016 à 2020**
 - **Statistiques de base (Points, passes, rebonds, contres...)**
 - **Statistiques Avancées (OWS, DWS, PER, etc..)**
-



```
graph LR; A[Pre-Processing] --> B[Modeling]; B --> C[Plotting]; C --> D[Déploiement]
```

Pre-Processing

Modeling

Plotting

Déploiement

Pre-Processing

- Normalisation des noms (Accents, points, Jr, etc..)
 - Récupérer la dernière équipe pour les joueurs transférés
 - Enlever les joueurs retraités depuis
 - Agréger les statistiques de base (Somme puis moyenne par match)
 - Agréger les statistiques avancées en pondérant par match joués
 - Enlever les joueurs avec un échantillon de jeu trop faible (en minutes et en Matches joués)
 - Ramener sur 36 minutes
 - Scaler sur [0, 1]
 - Scrapping pour récupérer la taille de joueurs
-

Modeling

- **Score Silhouette**
- **DBSCAN**
- **FCM**
- **Matrice de Distance**



Plotting

- Polygone
- Heat Matrix



Déploiement

- Streamlit



Conclusion

- Data-Centric plutôt que Model-Centric => Importance du Pre-Processing et de la qualité des données
 - Importance de la Visualisation de données
 - Projet transversal (Scrapping, pré-processing, modeling, plotting, déploiement)
-

Merci!

