



SCI 1402

Projet en science des données

PROJET

- Remplissez soigneusement les informations ci-dessous.
- Commencez votre travail à la page suivante.
- Sauvegardez votre travail de cette façon :
SCI1402-RapportProjet_Votrenom.docx
- Utilisez l'outil de dépôt des travaux (accessible par votre portail étudiant MaTÉLUQ) pour acheminer votre travail à la personne responsable de votre encadrement.

Votre prénom et nom : Kouadio Eden Elie Herve BOUSSON

Votre numéro d'étudiant : 24124235

Trimestre d'inscription : ETE

Date d'envoi : 2025-08-13

Étape 3 : Rapport final

Titre du projet

Prévision des performances offensives des joueurs NBA pour la saison 2024-2025 à partir des données de la saison 2023-2024

Nom du projet sélectionné

Analyse et modélisation des performances des joueurs NBA – Saison 2023-2024

Objectif général du projet

Construire un modèle de régression basé sur les statistiques NBA de la saison 2023-2024 afin de prévoir la moyenne de points par match (PTS/Game) des joueurs pour la saison suivante (2024-2025).

1. Introduction

Le basketball professionnel génère d'énormes volumes de données statistiques qui permettent d'analyser les performances des joueurs. Ce projet a pour objectif de prédire le nombre moyen de points par match pour la saison 2024-2025 à partir des données statistiques des playoffs de la saison 2023-2024, en utilisant des techniques de modélisation statistique en Python.

Les résultats permettront d'identifier les joueurs ayant un potentiel offensif élevé et d'offrir une base objective pour les analyses sportives et stratégiques.

2. Description des données

Source : Kaggle – 2023-2024 NBA Player Stats - Playoffs.csv

Type de données : Tableau CSV séparé par « ; »

Nombre initial d'observations : 241 joueurs

Variables explicatives sélectionnées : MIN, FG, FGA, FG_PCT, TRB, AST, STL, BLK, TOV

Variable cible : PTS

3. Nettoyage et préparation des données

1. Chargement du fichier CSV avec pandas (séparateur « ; »)
2. Suppression des valeurs manquantes
3. Filtrage des joueurs ayant disputé moins de 6 matchs

4. Sélection et renommage des colonnes pertinentes
5. Vérification de la cohérence des valeurs numériques

4. Méthodologie

1. Découpage des données : 80% apprentissage, 20% test
2. Modèle utilisé : Régression linéaire multiple (LinearRegression de scikit-learn)
3. Évaluation avec : R^2 , RMSE, MAE

5. Résultats

Indicateur	Résultat
R^2	0.97
RMSE	1.09
MAE	0.84

6. Tableau des prédictions pour 10 joueurs clés

Joueur	Points prédits
Jalen Brunson	32.25
Tyrese Maxey	29.51
Shai Gilgeous-Alexander	29.48
Nikola Jokić	29.01
Luka Dončić	28.86
Joel Embiid	28.69
Donovan Mitchell	28.66
Paolo Banchero	27.80
Anthony Edwards	26.52
Khris Middleton	25.13

7. Interprétation

Les variables les plus influentes sont : FG, MIN et FG_PCT. Un joueur ayant un haut pourcentage de réussite et plus de minutes jouées est fortement associé à un score plus élevé. La corrélation avec les passes (AST) est aussi notable.

8. Recommandations

- Tester une régularisation (Ridge ou Lasso) pour plus de robustesse
- Ajouter des variables contextuelles (domicile/extérieur, position)
- Étendre l'analyse à plusieurs saisons

9. Conclusion

Le modèle de régression linéaire multiple a obtenu une excellente performance ($R^2 = 0.97$). Il est fiable pour estimer les performances offensives des joueurs NBA et pourrait être amélioré avec plus de données et des modèles plus complexes.

1. Introduction

Le basketball professionnel génère d'énormes volumes de données statistiques qui permettent d'analyser les performances des joueurs. Ce projet a pour objectif de prédire le nombre moyen de points par match pour la saison 2024-2025 à partir des données statistiques des playoffs de la saison 2023-2024, en utilisant des techniques de modélisation statistique en Python.

Les résultats permettront d'identifier les joueurs ayant un potentiel offensif élevé et d'offrir une base objective pour les analyses sportives et stratégiques.

2. Description des données

Source : Kaggle – 2023-2024 NBA Player Stats - Playoffs.csv

Type de données : Tableau CSV séparé par « ; »

Nombre initial d'observations : 241 joueurs

Variables explicatives sélectionnées : MIN, FG, FGA, FG_PCT, TRB, AST, STL, BLK, TOV

Variable cible : PTS

3. Nettoyage et préparation des données

1. Chargement du fichier CSV avec pandas (séparateur « ; »)
2. Suppression des valeurs manquantes
3. Filtrage des joueurs ayant disputé moins de 6 matchs
4. Sélection et renommage des colonnes pertinentes
5. Vérification de la cohérence des valeurs numériques

4. Méthodologie

1. Découpage des données : 80% apprentissage, 20% test
2. Modèle utilisé : Régression linéaire multiple (LinearRegression de scikit-learn)
3. Évaluation avec : R^2 , RMSE, MAE

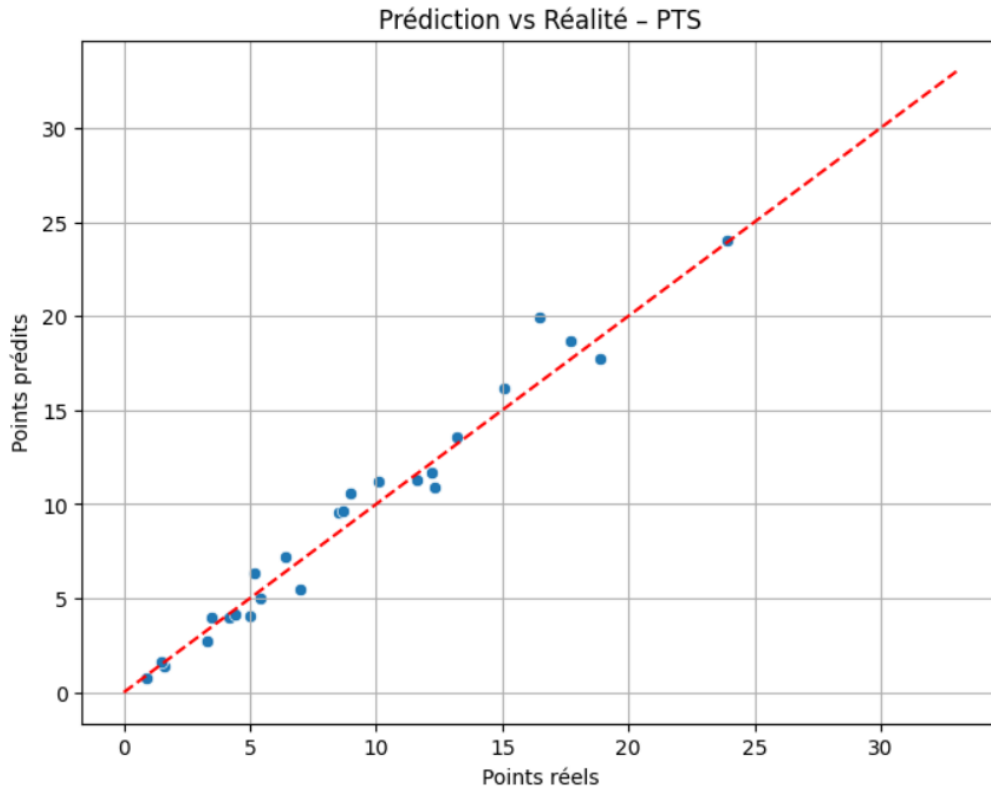
5. Résultats

Indicateur	Résultat
R^2	0.97
RMSE	1.09

MAE	0.84
-----	------

6. Tableau des prédictions pour 10 joueurs clés

Joueur	Points prédits
Jalen Brunson	32.25
Tyrese Maxey	29.51
Shai Gilgeous-Alexander	29.48
Nikola Jokić	29.01
Luka Dončić	28.86
Joel Embiid	28.69
Donovan Mitchell	28.66
Paolo Banchero	27.80
Anthony Edwards	26.52
Khris Middleton	25.13



7. Interprétation

Les variables les plus influentes sont : FG, MIN et FG_PCT. Un joueur ayant un haut pourcentage de réussite et plus de minutes jouées est fortement associé à un score plus élevé. La corrélation avec les passes (AST) est aussi notable.

8. Recommandations

- Tester une régularisation (Ridge ou Lasso) pour plus de robustesse
- Ajouter des variables contextuelles (domicile/extérieur, position)
- Étendre l'analyse à plusieurs saisons

9. Conclusion

Le modèle de régression linéaire multiple a obtenu une excellente performance ($R^2 = 0.97$). Il est fiable pour estimer les performances offensives des joueurs NBA et pourrait être amélioré avec plus de données et des modèles plus complexes.