Machine Learning
Finals Project

אריק טטייבסקי 208997056
רועי משולם 315635649

Data Description

NBA Salaries: Hoops Fortune (2020-2025)

https://www.kaggle.com/datasets/omarsobhy14/nba-players-salaries?resource=download

The "NBA Player Salaries 2023-2025" dataset is a treasure trove of financial insights in the basketball world. It features detailed records of player earnings for each season. We've also manually added each player's NBA 2K rating

Each object in our data contains

Player ID - *unique* integer

Player Full Name - string

Salary 22-23 - integer

Salary 23-24 - integer

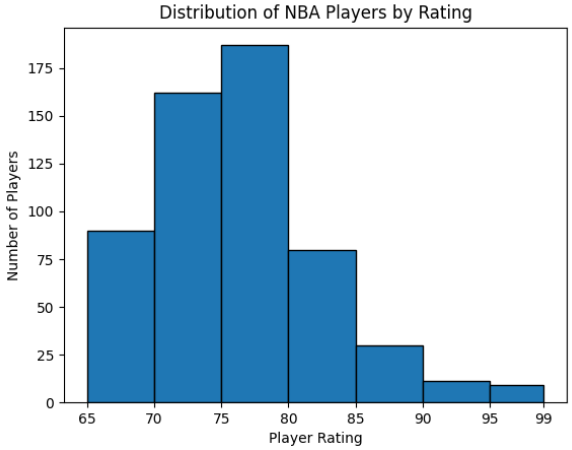
Salary 24-25 - integer

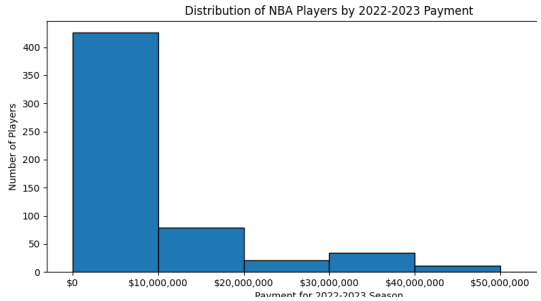
NBA 2K Rating - integer in the range of 60-99

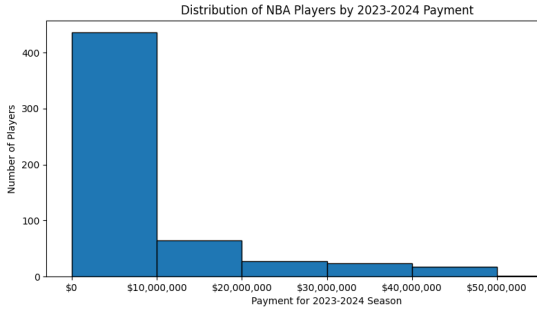
Partial screenshot from the csv file:

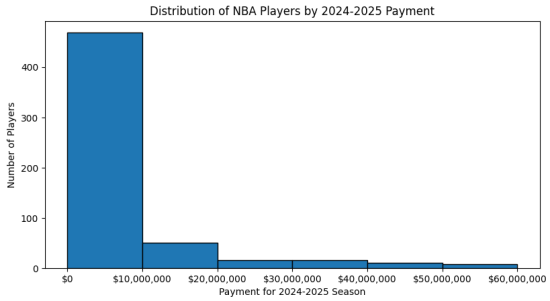
Player Id	Player Name	2022/2023	2023/2024	2024/2025	2K Rating
1	Stephen Curry	$48,070,014	$51,915,615	$55,761,217	96
2	John Wall	$47,345,760	$0	$0	79
3	Russell Westbrook	$47,080,179	$0	$0	81
4	LeBron James	$44,474,988	$46,698,737	$50,434,636	97
5	Kevin Durant	$44,119,845	$47,649,433	$51,179,020	96
6	Bradley Beal	$43,279,250	$46,741,590	$50,203,930	87
7	Paul George	$42,492,492	$45,640,084	$48,787,676	89
8	Kawhi Leonard	$42,492,492	$45,640,084	$48,787,676	92
9	Giannis Antetokounmpo	$42,492,492	$45,640,084	$48,787,676	97
10	Damian Lillard	$42,492,492	$45,640,084	$48,787,676	95
11	Klay Thompson	$40,600,080	$43,219,440	$0	86
12	Kyrie Irving	$38,917,057	$0	$0	91
13	Rudy Gobert	$38,172,414	$41,000,000	$43,827,586	84
14	Khris Middleton	$37,984,276	$40,396,552	$0	86
15	Anthony Davis	$37,980,720	$40,600,080	$43,219,440	94
16	Jimmy Butler	$37,653,300	$45,183,960	$48,798,677	93
17	Tobias Harris	$37,633,050	$39,270,150	$0	81
18	Kemba Walker	$37,281,261	$0	$0	76
19	Trae Young	$37,096,500	$40,064,220	$43,031,940	89
20	Zach LaVine	$37,096,500	$40,064,220	$43,031,940	87
21	Luka Doncic	$37,096,500	$40,064,220	$43,031,940	97
22	Ben Simmons	$35,448,672	$37,893,408	$40,338,144	78
23	Pascal Siakam	$35,448,672	$37,893,408	$0	87

Data Distribution









שאלה שנרצה לענות עליה:

אנחנו רוצים לדעת אם על ידי לימוד המשכורות של כל שנה נוכל לחזות את דירוג השחקנים החדשים הנכנסים ל-NBA

גישה

חילקנו את הנתונים באופן אקראי ב-50% לאימון ו-50% כדי לבדוק ולנסות לסווג אותם באמצעות האלגוריתמים הבאים

SVM

RANDOM FOREST

ADABOOST

K-NEAREST-NEIGHBORS

בעיות שאנחנו צופים עם הדאטה:

משכורת גבוהה וביצועים לא טובים

Ben Simmons

22	Ben Simmons	$35,448,672	$37,893,4(	$40,338,1(	78

John Wall

2	John Wall	$47,345,760	$0	$0	79

Kemba Walker

18	Kemba Walker	$37,281,261	$0	$0	76

משכורת נמוכה וביצועים מאד טובים

Desmond Bane

369	Desmond Bane	$2,130,240	$3,845,08	$5,767,62	85

LaMelo Ball

169	LaMelo Ball	$8,623,920	$10,900,6	$14,301,6	86

Results - SVM

Support Vector Machine (SVM)

Mean Squared Error: 14.649878751887888

Sample screenshot

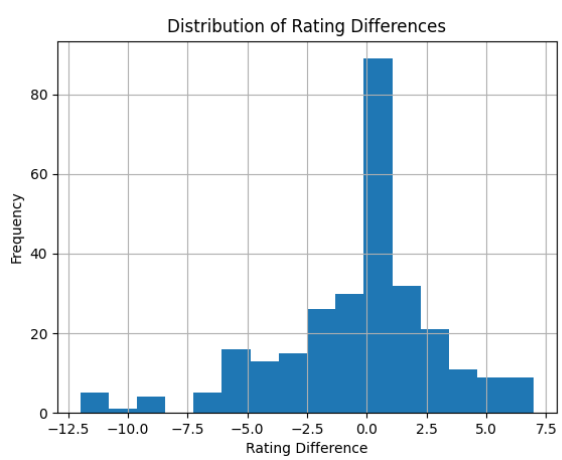
```
Player: Bojan Bogdanovic
  Predicted Rating = 81.97
  Actual Rating = 82.0

Player: RJ Barrett
  Predicted Rating = 83.42
  Actual Rating = 82.0

Player: Ron Harper Jr
  Predicted Rating = 71.12
  Actual Rating = 69.0

Player: Shaquille Harrison
  Predicted Rating = 71.00
  Actual Rating = 69.0

Player: Walker Kessler
  Predicted Rating = 73.00
  Actual Rating = 83.0
```



Results - Random Forest

Random Forest

Mean Squared Error: 10.86627822690605

Sample screenshot

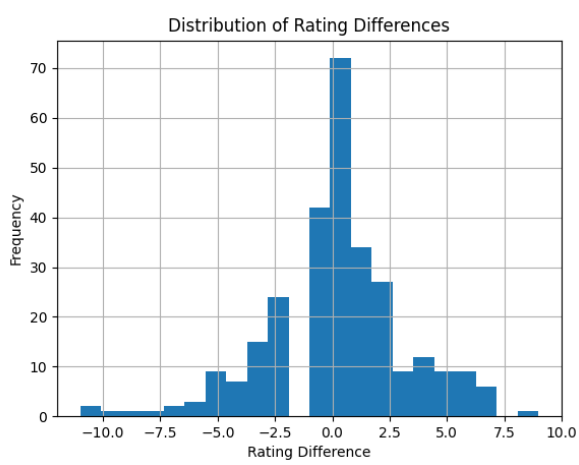
```
Player: Bojan Bogdanovic
  Predicted Rating = 82.19
  Actual Rating = 82.0

Player: RJ Barrett
  Predicted Rating = 83.18
  Actual Rating = 82.0

Player: Ron Harper Jr
  Predicted Rating = 69.16
  Actual Rating = 69.0

Player: Shaquille Harrison
  Predicted Rating = 68.78
  Actual Rating = 69.0

Player: Walker Kessler
  Predicted Rating = 72.90
  Actual Rating = 83.0
```



Results - Adaboost

Adaboost

Mean Squared Error: 10.59777273779598

Sample screenshot

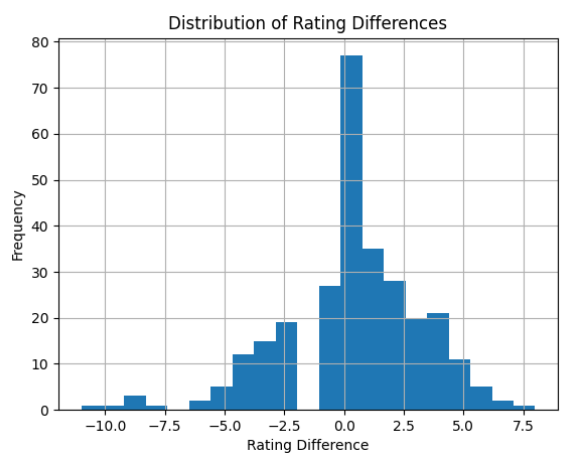
```
Player: Bojan Bogdanovic
  Predicted Rating = 80.66
  Actual Rating = 82.0

Player: RJ Barrett
  Predicted Rating = 86.11
  Actual Rating = 82.0

Player: Ron Harper Jr
  Predicted Rating = 69.14
  Actual Rating = 69.0

Player: Shaquille Harrison
  Predicted Rating = 69.34
  Actual Rating = 69.0

Player: Walker Kessler
  Predicted Rating = 73.90
  Actual Rating = 83.0
```



Results - KNN

K-Nearest-Neighbors (KNN)

Mean Squared Error: 9.924475524475524

Sample screenshot

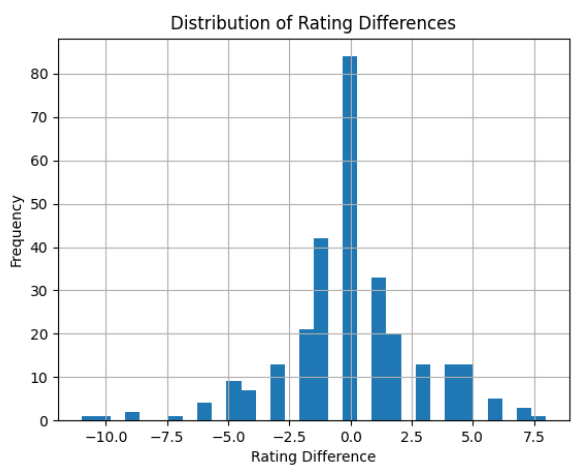
```
Player: Bojan Bogdanovic
  Predicted Rating = 83.00
  Actual Rating = 82.0

Player: RJ Barrett
  Predicted Rating = 83.40
  Actual Rating = 82.0

Player: Ron Harper Jr
  Predicted Rating = 68.80
  Actual Rating = 69.0

Player: Shaquille Harrison
  Predicted Rating = 69.00
  Actual Rating = 69.0

Player: Walker Kessler
  Predicted Rating = 73.60
  Actual Rating = 83.0
```



Result - Analysis

עשינו השוואה בין ערכי הMSE של ארבעת האלגוריתמים

התוצאות הן:

$\text{SVM} = 14.65$, $\text{Random Forest} = 10.87$, $\text{AdaBoost} = 10.60$, $\text{KNN} = 9.92$

עשינו השוואה בין ערכי הMSE של ארבעת האלגוריתמים

ערכי MSE מציינים את ההבדל הממוצע בריבוע בין הדירוג החזוי לדירוג בפועל. MSE נמוך יותר מסמל דיוק חיזוי טוב יותר.

ההבדל בין ערכי הMSE בין האלגוריתמים השונים:

<u>גישה אלגוריתמית:</u> כל אלגוריתם משתמש במתודולוגיה ייחודית לניבוי הדירוגים. SVM שואפת למצוא מישור המפריד בין הנתונים, Random Forest משלב עצי החלטה מרובים, AdaBoost מתמקדת בהרכבת לומדים חלשים, ו-KNN מסתמכת על הקרבה של נקודות נתונים.

<u>מורכבות המודל</u>: לאלגוריתמים שונים יש רמות שונות של מורכבות המודל. ל-SVM, Random Forest, AdaBoost ו-KNN יש הנחות יסוד שונות, גבולות החלטה וטיפול בקשרים לא ליניאריים. השונות הללו יכולות להשפיע על היכולת שלהם לתפוס את המורכבויות של מערך הנתונים של ה-NBA בצורה מדויקת.

<u>חשיבות התכונה</u>: התכונות (המשכורות) שנבחרו המשמשות לחיזוי יכולות להשפיע על ביצועי האלגוריתמים. בעוד שכל האלגוריתמים השתמשו באותן תכונות, הם עשויים להקצות רמות שונות של חשיבות לכל תכונה, וכתוצאה מכך ערכי MSE מגוונים.

למה דווקא KNN יצר לנו את הMSE הכי טוב?

KNN השיג את ה-MSE הנמוך ביותר מבין ארבעת האלגוריתמים, מה שמצביע על דיוק חיזוי טוב יותר יחסית. ישנן כמה סיבות לכך ש-KNN עשויה להתעלות על האלגוריתמים האחרים:

קרבה מקומית: KNN מנצל את הקרבה המקומית של נקודות נתונים כדי לבצע תחזיות. במערך הנתונים של ה-NBA, שחקנים עם משכורות דומות עשויים להציג דירוגים דומים של 2K, מה שהופך את הגישה המקומית של KNN ליעילה בלכידת מערכות יחסים אלו.

שיפור הדיוק של האלגוריתמים

<u>תכונות נוספות</u>: גורמים כגון סטטיסטיקות ביצועי שחקנים, דינמיקה של צוות או חווית שחקן עשויים לספק תובנות חשובות ולשפר את כוח הניבוי של המודלים. כמו כן אנו יכולים לקחת פרמטרים נוספים כגון גובה, משקל וגיל.

<u>איכות וכמות הנתונים</u>: ודא שמערך הנתונים נקי ומעודכן (חוזים חדשים).

Experiment 2

We will try to improve each of the algorithms to generate a better mse:

SVM

ניסינו את שלושת הפונקציות שניתן להכניס לSVM כפרמטר

RBF - 14.649

Sigmoid - 267.163

Poly - 26.001

Random Forest

שינינו את המשתנים של n_estimators, max_depth כדי למזער את ה-mse וגילינו שכאשר n_estimators=65, max_depth=3 אנו יוצרים mse = 9.676 (המקור הוא 11.309).

n_estimators: פרמטר זה מציין את מספר העצים ביער האקראי.

max_depth: העומק המרבי של כל עץ ביער האקראי.

Adaboost

שינינו את המשתנה learning_rate כדי למזער את ה-mse וגילינו שכאשר learning_rate=0.16 אנו יוצרים mse = 9.759 (המקור הוא 10.357).

learning_rate: פרמטר זה הוא המשקל המוחל על כל רגרסור בכל איטרציה מגבירה. שיעור למידה גבוה יותר מגדיל את התרומה של כל רגרסור.

KNN

שינינו את המשתנה N-Neighbors כדי למזער את ה-mse וגילינו שכאשר n_neighbors=7 אנחנו יוצרים mse = 9.524 (המקור הוא 9.924).

n_neighbors: פרמטר זה מציין את מספר השכנים שאנו רוצים.

Summary Experiment 2

SVM:

Original MSE: 14.649

Optimized MSE: 14.649

Random-Forest:

Original MSE: 11.309

Optimized MSE: 9.676

AdaBoost:

Original MSE: 10.357

Optimized MSE: 9.759

KNN:

Original MSE: 9.924

Optimized MSE: 9.524

Experiment 3

נשתמש בשיפורים שעשינו בניסוי 2 והפעם נחלק את הDATA ל80-20

Summary Experiment 3

<u>SVM:</u>
Experiment 2 MSE: 14.649
Optimized MSE: 12.228
<u>Random-Forest:</u>
Experiment 2 MSE: 9.676
Optimized MSE: 9.522
<u>AdaBoost:</u>
Experiment 2 MSE: 9.759
Optimized MSE: 8.973
<u>KNN:</u>
Experiment 2 MSE: 9.524
Optimized MSE: 8.375

Predict new incoming players to the NBA ratings:

Arik Tatievski, 15,000,000 , 20,000,000 , 25,000,000

Roi Meshulam, 2,500,000 , 4,000,000 , 10,000,000

SVM

```
Player: Arik Tatievski
  Predicted Rating = 82.85

Player: Roi Meshulam
  Predicted Rating = 76.30
```

Random Forest

```
Player: Arik Tatievski
  Predicted Rating = 81.50

Player: Roi Meshulam
  Predicted Rating = 78.16
```

AdaBoost

```
Player: Arik Tatievski
  Predicted Rating = 81.26

Player: Roi Meshulam
  Predicted Rating = 78.66
```

KNN

```
Player: Arik Tatievski
  Predicted Rating = 82.60

Player: Roi Meshulam
  Predicted Rating = 76.80
```