

TEKNOKENT YZ PROJESİ

ns, RBI, Walks, Years, CAtBat, CHits, CHmRun, CRuns, CRBI, CWalks, League, Division, P
, 293, 66, 1, 30, 29, 14, A, E, 446, 33, 20, NA, A
4, 3449, 835, 69, 321, 414, 375, N, W, 632, 43, 10, 475, N
, 3, 1624, 457, 63, 224, 266, 263, A, W, 880, 82, 14, 480, A
, 11, 5628, 1575, 225, 828, 838, 354, N, E, 200, 11, 3, 500, N
2, 396, 101, 12, 48, 46, 33, N, E, 805, 40, 4, 91.5, N
11, 4408, 1133, 19, 501, 336, 194, A, W, 282, 421, 25, 750, A
214, 42, 1, 30, 9, 24, N, E, 76, 127, 7, 70, A
509, 108, 0, 41, 37, 12, A, W, 121, 283, 9, 100, A
341, 86, 6, 32, 34, 8, N, W, 143, 290, 19, 75, N
13, 5206, 1332, 253, 784, 890, 866, A, E, 0, 0, 0, 1100, A
9, 10, 4631, 1300, 90, 702, 504, 488, A, E, 238, 445, 22, 517.143, A
, 1876, 467, 15, 192, 186, 161, N, W, 304, 45, 11, 512.5, N
, 4, 1512, 392, 41, 205, 204, 203, N, E, 211, 11, 7, 550, N
, 1941, 510, 4, 309, 103, 207, A, E, 121, 151, 6, 700, A
3, 3231, 825, 36, 376, 290, 238, N, E, 80, 45, 8, 240, N
, 201, 42, 3, 20, 16, 11, A, W, 118, 0, 0, NA, A
, 15, 8068, 2273, 177, 1045, 993, 732, N, W, 105, 290, 10, 775, N
479, 102, 5, 65, 23, 39, A, W, 102, 177, 16, 175, A
12, 5233, 1478, 100, 643, 658, 653, A, W, 912, 88, 9, NA, A
727, 180, 24, 67, 82, 56, N, W, 202, 22, 2, 135, N
1, 413, 92, 16, 72, 48, 65, N, E, 280, 9, 5, 100, N
1, 426, 109, 3, 55, 43, 62, A, W, 361, 22, 2, 115, N
26, 2, 9, 9, 3, A, W, 812, 84, 11, NA, A
, 6, 1924, 489, 67, 242, 251, 240, N, W, 518, 55, 3, 600, N
0, 18, 8424, 2464, 164, 1008, 1072, 402, A, E, 1067, 157, 14, 776.667, A
6, 2695, 747, 17, 442, 198, 317, A, E, 434, 9, 3, 765, A
, 1931, 491, 13, 291, 108, 180, N, E, 222, 3, 3, 708.333, N
, 10, 2331, 604, 61, 246, 327, 166, N, W, 732, 83, 13, 750, N
5, 2308, 633, 32, 349, 182, 308, N, W, 262, 329, 16, 625, N
, 14, 5201, 1382, 166, 763, 734, 784, A, W, 267, 5, 3, 900, A
7, 6890, 1833, 224, 1033, 864, 1087, A, W, 127, 221, 7, NA, A
3, 591, 149, 8, 80, 46, 31, N, W, 226, 7, 4, 110, N

PROJE HAKKINDA

Bu projede, Hitters veri setinde yer alan bağımsız değişkenler kullanılarak bağımlı değişken olan salary değerinin tahmin edilmesi hedeflenmiştir. Çalışma kapsamında veri analizi ve makine öğrenmesi tekniklerinden faydalanılmış olup, aşağıdaki Python kütüphaneleri kullanılmıştır: Pandas (veri manipülasyonu ve analizi), NumPy (sayısal hesaplama ve veri işleme), Matplotlib ve Seaborn (veri görselleştirme), Scikit-learn (makine öğrenmesi modelleri ve değerlendirme metrikleri). Projede, veri seti detaylı bir şekilde analiz edilmiş, bağımsız değişkenler ile salary değişkeni arasındaki ilişkiler incelenmiş ve uygun modeller geliştirilerek maaş tahminleri gerçekleştirılmıştır.



METOT VE YÖNTEMLER

Projede, maaş tahmini için Random Forest modeli kullanılmıştır. Modelin performansını artırmak amacıyla, modelin hiperparametrelerinden n_estimators, max_depth, min_samples_split ve max_features değerleri, k katlı çapraz doğrulama yöntemiyle optimize edilmiştir. Bu süreçte, en uygun hiperparametre değerleri belirlenerek modelin ayarları (tuning) gerçekleştirilmiştir. Modelin performansını değerlendirmek için Mean Squared Error (MSE), Root Mean Squared Error (RMSE) ve Mean Absolute Error (MAE) gibi performans metriklerinden faydalanyanmıştır.

```
klearn.metrics import mean_squared_error,mean_absolute_error
klearn.ensemble import RandomForestRegressor
klearn.model_selection import train_test_split,GridSearchCV
pandas as pd
numpy as np
matplotlib.pyplot as plt
seaborn as sns

pd.read_csv("/content/Hitters.csv")
datas.dropna()
.get_dummies(datas[["League","NewLeague","Division"]])
s["Salary"]
as.drop(["Salary","League","NewLeague","Division"],axis=1).astype("float64")
oncat([x_,dms],axis=1)
n,x_test,y_train,y_test=train_test_split(x,y,test_size=0.25,random_state=99)

RandomForestRegressor()
ramps={
    "n_estimators": [50,100,150],
    "max_depth": [10,20,30],
    "min_samples_split": [2,4],
    "max_features": [1,2,3]

}=GridSearchCV(RFR,RFR_params,cv=2,n_jobs=-1,verbose=2)
.fit(x_train,y_train)
mators=RFR_cv.best_params_["n_estimators"]
pth=RFR_cv.best_params_["max_depth"]
mples_split=RFR_cv.best_params_["min_samples_split"]
atures=RFR_cv.best_params_["max_features"]
ned=RandomForestRegressor(n_estimators=n_estimators,max_depth=max_depth,min_samples_split=tuned_params["min_samples_split"],max_features=tuned_params["max_features"])
ned.fit(x_train,y_train)
t=RFR_tuned.predict(x_test)
an_squared_error(y_test,predict)
n_sqrd(MSE)
```



VERİ SETİ

Veri seti, AtBat, Hits, HmRun, Runs, RBI, Walks, Years, CAtBat, CHits, CHmRun, CRuns, CRBI, CWalks, League, Division, PutOuts, Assists, Errors ve NewLeague bağımsız değişkenleri ile Salary bağımlı değişkenini içermektedir. Veri seti üzerinde çeşitli düzenlemeler ve farklı değişiklikler yapılarak, regresyon modellerine uygun hale getirilmiştir. Bu yapı, bağımsız değişkenler kullanılarak Salary tahmini için etkili analizler yapılmasına olanak sağlamaktadır.

Modelin tahmin performansı makul düzeyde görünüyor; ancak RMSE ve MAE değerleri, özellikle maaş tahminleri için daha düşük olsaydı, model daha iyi sonuçlar veriyor olurdu.

Metrik sonuçları

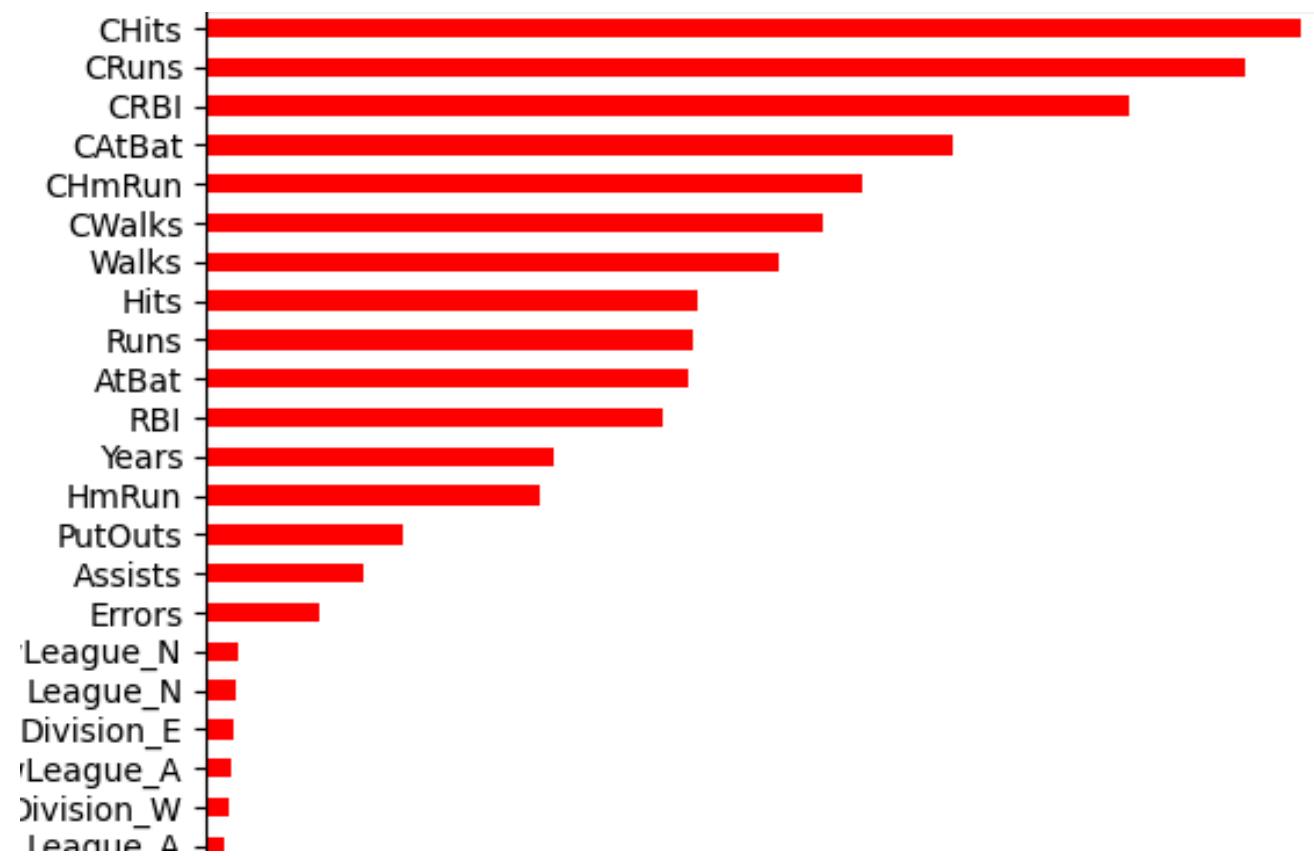
MSE: 133537.32128020521

RMSE: 365.427586917306

MAE: 220.7185764708339



GRAFİKLER

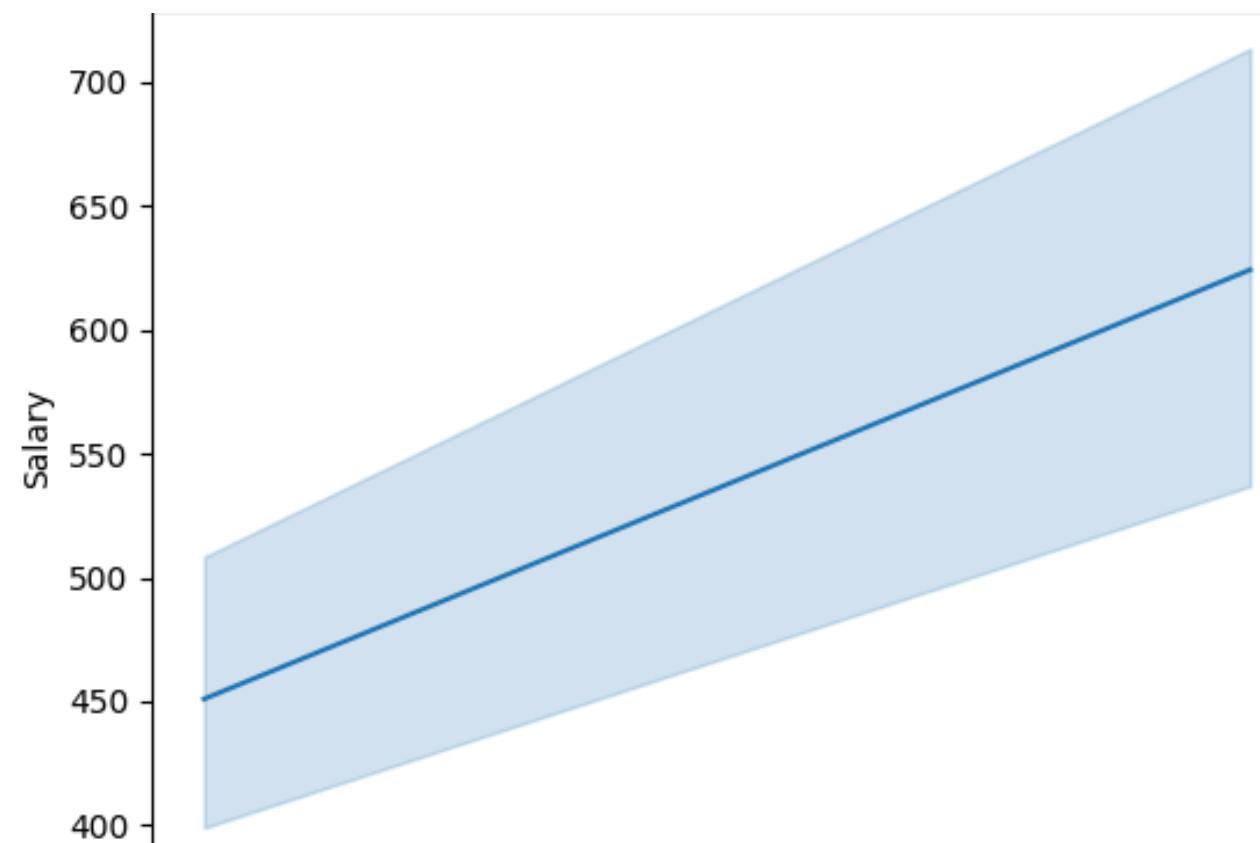


01

Değişkenlerin Önem Sırası

Bu grafikte, bağımsız değişkenlerin hedef değişken (Salary) üzerindeki önem sırası görselleştirilmiştir. CHits, CRuns, ve CRBI gibi değişkenler en yüksek öneme sahiptir. Bu, bir oyuncunun kariyerindeki toplam hit sayısının (CHits), toplam koşularının (CRuns), ve toplam RBI değerinin (Runs Batted In) maaşını tahmin etmekte en etkili değişkenler olduğunu gösterir.

Daha az önemli olan değişkenler arasında League_N, Division_E, ve League_A yer almaktadır. Bu, oyuncunun bulunduğu ligin veya bölümün maaş üzerinde nispeten daha az etkisi olduğunu göstermektedir.



02

Division ve Salary İlişkisi

Bu grafikte, Division (lig bölümü) ve Salary arasındaki ilişki bir regresyon doğrusu ile gösterilmiştir.

- Division_W (Bati) bölgesindeki oyuncuların maaşları daha düşük, Division_E (Doğu) bölgesindeki oyuncuların maaşları ise daha yüksektir.
- Aradaki eğilim doğrusal bir ilişkiyi göstermektedir; bu, Division değişkeninin Salary tahmininde etkili olabileceğini, ancak katkısının sınırlı olduğunu da ima eder (ilk grafikten de bu değişkenin düşük öneme sahip olduğu görülmüştür).

Özetle, bir oyuncunun bulunduğu bölge maaş seviyelerini etkiler, ancak bu etki diğer performans göstergeleri (CHits, CRuns, vb.) kadar belirgin değildir.



DİNLEDİĞİNİZ İÇİN TEŞEKKÜRLER