

Makine Öğrenmesi Kullanarak FIFA 23 Oyunundaki Futbolcuların Piyasa Değerinin Tahmin Edilmesi

Muhammet Semih KELEŞ
İzmir Kâtip Çelebi Üniversitesi
İzmir, Türkiye
semihkeles1997@hotmail.com

I. GİRİŞ (INTRODUCTION)

Oyun, kendiliğinden ortaya çıkan hedefi olmayan, mutluluk getiren serbest bir aktivite olarak tanımlanmaktadır. Oyunlar kendi içerisinde motor ve fiziksel oyunlar, sosyal oyunlar, yapısal oyunlar, kurallı oyunlardır. Oyunlar her açıdan farklı hareket ve düşünme sistemini, alışagelmış davranış şekillerinin aksine başkalarının davranışlarına ve durumun şartlarına göre kurgulanmış düşünce ve hareket biçimleri olarak da değerlendirilirler (Wood & Attfield, 2005) [1]. Futbol da belirli kurallar dahilinde oynanan ve dünyada en çok sevilen ve oynanan sporlardan bir tanesidir. Ancak dünyada çok az insanın profesyonel futbolcu olabilmesi birçok insanı hayal kırıklığına uğratmaktadır. Bu sebeple de teknolojinin gelişmesiyle beraber sporlar sanal ortamda oynanan oyunlar haline de gelmiş ve hem insanların ihtiyaçlarını karşılamış hem de eğlenceli vakit geçirmelerini sağlamıştır. Sanal ortamda da en çok tercih edilen sporlardan biri yine futbol olmuştur. Öyle ki her sene milyonlarca oyuncusu olduğu tahmin edilmektedir. Bu alanda Pro Evolution Soccer (PES) ve Fédération Internationale de Football Association (FIFA) olmak üzere çok önemli 2 şirket yer almaktadır. Bu şirketler dünyanın en büyük futbol oyun endüstrisinin başlarında gelmektedir. Futbol oyunlarında her futbolcunun farklı özellikleri bulunabilmekte ve yine her futbolcunun yetenek değerlendirmesi diğerlerinden farklı olabilmektedir. Örneğin dünyanın en iyi futbolcuları olarak görülen Cristiano Ronaldo ve Lionel Messi gibi futbolcular şut, hız, hızlanma, duran top, kafa vuruş yeteneği gibi birçok özellik değerlerinde diğer futbolculardan iyi olabilmektedir. Onları en iyisi yapan da bu özelliklerdir. Aynı şekilde bu futbolcuları takımlarında oynatmak isteyen takımlar ise bu futbolcuları alabilmek için futbolcunun mevcut takımına belli bonservis ücreti ödemek durumunda kalabilmektedirler. Bu ücretin belirlenmesinde futbolcunun genel yetenekleri, futbolcunun mevcut takımıyla kalan sözleşme süresi, futbolcunun takımının kendi ligindeki ve dünya çapındaki sıralaması, futbolcunun mevcut takımının katıldığı organizasyon (Şampiyonlar Ligi, UEFA Avrupa Ligi, UEFA Konferans Ligi gibi), futbolcunun yaşı, gelişme durumu, futbolcunun serbest kalma ücreti gibi birçok özellik etkili olmaktadır. Bu çalışmada da hem mevcut futbolcunun hem de yeni oluşturulan

futbolcunun özelliklerine bakılarak piyasa değerinin tahmin edilmesi amaçlanmaktadır. İlgili çalışmada 18540 örnek ve "ID", "Name", "FullName", "Age", "Height", "Weight", "PhotoUrl", "Nationality", "Overall", "Potential", "Growth", "TotalStats", "BaseStats", "Positions", "BestPosition", "Club", "ValueEUR", "WageEUR", "ReleaseClause", "ClubPosition", "ContractUntil", "ClubNumber", "ClubJoined", "OnLoad", "NationalTeam", "NationalPosition", "NationalNumber", "PreferredFoot", "IntReputation", "WeakFoot", "SkillMoves", "AttackingWorkRate", "DefensiveWorkRate", "PaceTotal", "ShootingTotal", "PassingTotal", "DribblingTotal", "DefendingTotal", "PhysicalityTotal", "Crossing", "Finishing", "HeadingAccuracy", "ShortPassing", "Volleys", "Dribbling", "Curve", "FKAccuracy", "LongPassing", "BallControl", "Acceleration", "SprintSpeed", "Agility", "Reactions", "Balance", "ShotPower", "Jumping", "Stamina", "Strength", "LongShots", "Aggression", "Interceptions", "Positioning", "Vision", "Penalties", "Composure", "Marking", "StandingTackle", "SlidingTackle", "GKDividing", "GKHandling", "GKKicking", "GKPositioning", "GKReflexes", "STRating", "LWRating", "LFRating", "CFRating", "RFRating", "RWRating", "CAMRating", "LMRating", "CMRating", "RMRating", "LWBRating", "CDMRating", "RWBRating", "LBRating", "CBRating", "RBRating", "GKRating", "Club_Overall" olmak üzere 91 öznitelik bulunmaktadır. Bu özniteliklerden en uygun olanları seçilmiştir. Makine Öğrenmesinin Denetimli Öğrenme (Supervised Learning) modeli, tekniklerinden ise futbolcunun piyasa değerinin tahmin edilmeye çalışılması ve bu değerın sayısal bir karşılığının olması sebebiyle regresyon kullanılmıştır. Makine öğrenme algoritmalarından ise en çok kullanılan supervised algoritmalarından biri olan KNN (K-Nearest Neighbour), Karar Ağaçları (Decision Tree), Lineer Regresyon (Linear Regression), Destek Vektör Makineleri (Support Vector Machine) ve Sinir Ağları (Neural Network) kullanılmıştır.

II. YÖNTEMLER (METHODS)

A. Veri Seti (Data Set)

İlgili çalışmada FIFA 23 oyununda aktif olarak bulunan 18540 örnek ve 91 öznitelik içeren veri seti

kullanılmıştır. 91 öznitelik arasından 'Age', 'Height', 'Weight', 'Nationality', 'Overall', 'Potential', 'Growth', 'TotalStats', 'BaseStats', 'Positions', 'BestPosition', 'Club', 'ClubPosition', 'ContractUntil', 'OnLoad', 'NationalTeam', 'PreferredFoot', 'IntReputation', 'WeakFoot', 'SkillMoves', 'AttackingWorkRate', 'DefensiveWorkRate', 'PaceTotal', 'ShootingTotal', 'PassingTotal', 'DribblingTotal', 'DefendingTotal', 'PhysicalityTotal', 'Crossing', 'Finishing', 'HeadingAccuracy', 'ShortPassing', 'Volleys', 'Dribbling', 'Curve', 'FKAccuracy', 'LongPassing', 'BallControl', 'Acceleration', 'SprintSpeed', 'Agility', 'Reactions', 'Balance', 'ShotPower', 'Jumping', 'Stamina', 'Strength', 'LongShots', 'Aggression', 'Interceptions', 'Positioning', 'Vision', 'Penalties', 'Composure', 'Marking', 'StandingTackle', 'SlidingTackle', 'GKDividing', 'GKHandling', 'GK Kicking', 'GK Positioning', 'GK Reflexes', 'STRating', 'LWRating', 'LFRating', 'CFRating', 'RFRating', 'RWRating', 'CAMRating', 'LMRating', 'CMRating', 'RMRating', 'LWBRating', 'CDMRating', 'RWBRating', 'LBRating', 'CBRating', 'RBRating', 'GKRating', 'Club_Overall', olmak üzere 80 öznitelik seçilmiştir. Daha sonra geliştirilen fonksiyon yardımıyla ContractUntill ile günümüz tarihi çıkarılarak futbolcunun kalan kontrol yılı hesaplanmış ve buna bağlı olarak da ContractUntil özniteliği yerine Remaining_Contract özniteliği eklenmiştir. Bununla birlikte futbolcunun oynadığı takımın, tercih ettiği ayağın, kulübünde oynadığı pozisyonun, oynayabildiği pozisyonların, oynadığı en iyi pozisyonun ve oynadığı milli takımın ağırlıkları ayrı ayrı hesaplanmıştır. Bulunan değerler çok yüksek olduğundan (oyuncuların piyasa değerlerinin 100000 - 1000000000 civarı olduğu düşünüldüğünde) elde edilen ağırlık değerleri 000001 ile çarpılmıştır. Daha sonra bu değerler bağımlı değişkene (X) eklenmiştir.

B. Deneylerde Kullanılan Algoritmalar

K-En Yakın Komşu Algoritması (K-Nearest Neighbour KNN Algorithm)

Sınıflandırma ve regresyon işlemlerinde kullanılabilen ve sıklıkla tercih edilen algoritmalarından biridir. Özniteliklerin birbirlerine olan uzaklıklarının Öklid işlemi ile hesaplanması ile isminde de geçen K parametresi ile en yakın K sayıdaki örneğe bakarak sınıflandırma işleminin yapıldığı algoritmadır. [2] [3] İşlemin sonunda örneğimiz, K parametresi kadar örneğin içerisinde bulunulan sınıfa giriyor ise KNN algoritması çoğunluğun bulunduğu etiketi vermektedir. Örneğin K parametresi 3 verilmiş ve işlemin sonunda ulaşılmış örneklerin sınıf etiketleri "Başarılı", "Başarılı" ve "Başarısız" ise KNN algoritması etiketlemek istediğimiz örneği "Başarılı" etiketi ile etiketleyecektir. Özellikle sınıflandırma işlemlerinde genellikle K değeri tek sayı olarak seçilmektedir. Bunun sebebi işlemin sonunda elde edilecek etiketlerin yarı yarıya olma durumunu engellemektir. Örneğin K değerinin 4 seçilme durumunda ve işlemin sonunda "Başarılı", "Başarılı", "Başarısız" ve "Başarısız" etiketlerine sahip örneklerin

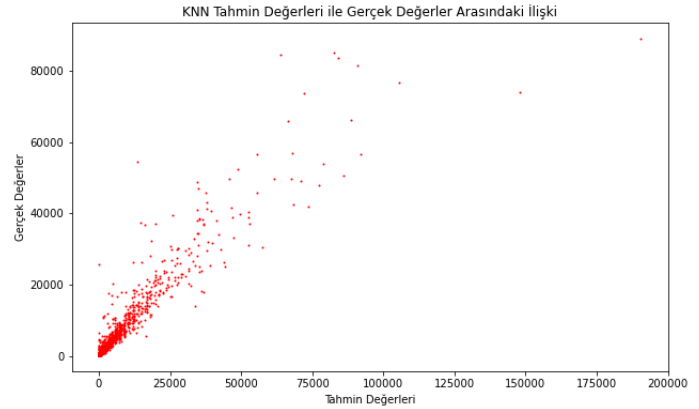
getirilmesi sonucunda KNN algoritması sağlıklı bir etiketleme işlemi yapamayacaktır. Regresyon işleminde de aynı işlemler yapıp K kadar örneğin ortalaması hesaplanmaktadır.

İlgili örnek için KNN algoritması hazırlanırken öncelikle uygun K değeri bulunmalıdır. Bunun için 1-50 arasındaki K değerleri için cross validation değeri 10 belirlenerek GridSearchCV algoritması uygulanmıştır. Uygulanan algoritma sonucunda en uygun K değeri 8 olarak bulunmuştur. İlgili sonuçlar ışığında KNN algoritması en uygun K değeri için eğitildiğinde sonuçlar *Tablo KNN1* 'deki gibidir.

Train Score	0.9131
Test Score	0.8685
Mean Absolute Error	619.1727
Mean Squared Error	8575458.8417
Root Mean Squared Error	2928.3884
R2 Score	0.8685

Tablo KNN1

İlgili sonuçlar ışığında oluşan KNN algoritmasının grafiği *Grafik KNN1* 'de gösterilmektedir.



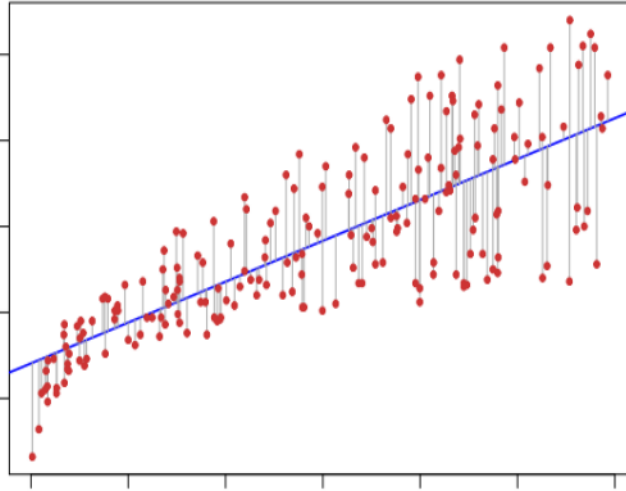
Grafik KNN1

Doğrusal Regresyon (Linear Regression)

"Basit doğrusal regresyon analizi (Neter, Wasserman, & Kutner, 1989), iki sürekli değişken arasındaki ilişkiyi nicelendiren istatistiksel bir tekniktir. Bu değişkenler, bağımlı değişken veya tahmin etmeye çalıştığımız değişken ve bağımsız veya tahmin edici değişken olarak ifade edilebilir" [4] Tahmin edilen değerler ile gerçek değerler arasındaki hata değerini minimumda tutmak üzerine çalışır. Formülü ise aşağıdaki gibidir.

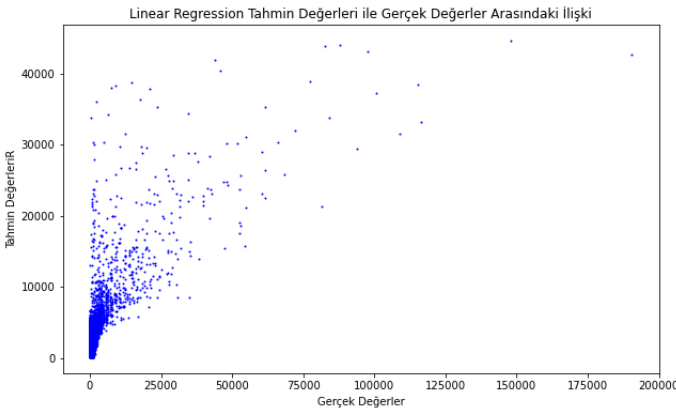
$$\min n \frac{\sum_{i=1}^n (y_i - \bar{y}_i)^2}{1}$$

Doğrusal Regresyon'un çalışma prensibi ise *Grafik LR1*'deki gibidir.



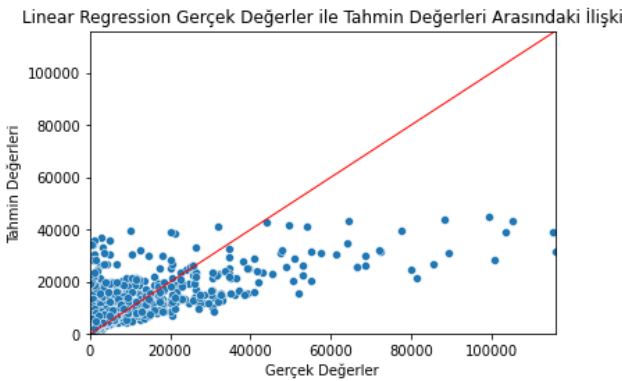
Grafik LR1 [4]

İlgili örnek için oluşturulan Doğrusal Regresyon için elde edilen grafik *Grafik LR2*'de belirtilmiştir.



Grafik LR2

Elde edilen grafik için çizilen doğru ise *Grafik LR3*'de belirtilmiştir.



Grafik LR3

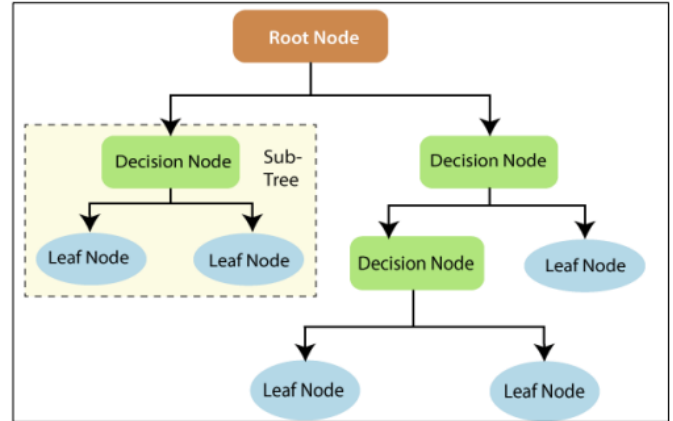
Uygulanan Doğrusal Regresyon için elde edilen sonuçlar ise *Tablo LR1*'de belirtilmektedir.

Train Score	0.5254
Test Score	0.5510
Mean Absolute Error	2388.2088
Mean Squared Error	28263782.2510
Root Mean Squared Error	5316.3692
R2 Score	0.5254

Tablo LR1

Karar Ağacı (Decision Tree)

Karar Ağacı isminden de anlaşılacağı üzere ağaç biçiminde bir yapı oluşturularak karar verme işleminin yapıldığı bir yapıdır. Karar ağaçlarında karar düğümleri (decision nodes) ve yaprak düğümleri (leaf nodes) bulunmaktadır. Karar düğümleri; karar verme, sınıflandırma yapma, tahmin etme gibi işlemleri gerçekleştirirken yaprak düğümleri ise bu kararların tutulduğu yerdir. Ağacın başı kök düğüm (root node) olarak isimlendirilir ve karara bağlanmak için kök düğümden yaprak düğümlere doğru belli bir yol izlenmektedir. [5]



Karar Ağacı (Decision Tree) [6]

İlgili örnek için uygulanan diğer algoritma olan Karar Ağacı Algoritması için Max Depth değeri 5 olarak belirlenmiştir. Oluşturulan algoritmadan çıkarın karar ağacı modeli *Grafik DT1*'de gösterilmektedir.



Grafik DT1

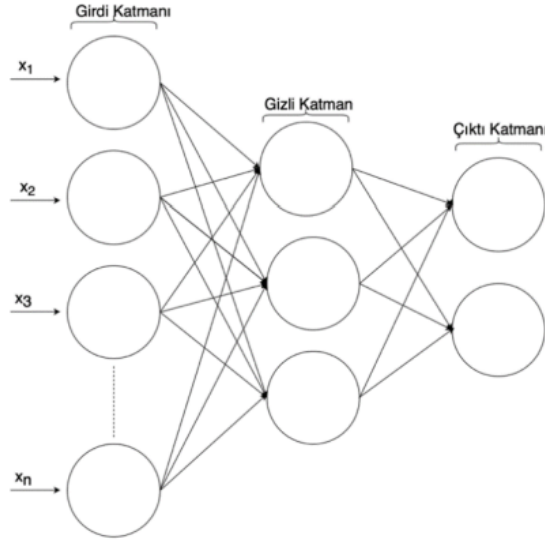
İlgili sonuçlar ışığında elde edilen sonuçlar ise *Tablo DT1*'de gösterilmektedir.

Train Score	0.9443
Test Score	0.8975
Mean Absolute Error	695.3842
Mean Squared Error	6091343.2607
Root Mean Squared Error	2468.0646
R2 Score	0.8975

Tablo DT1

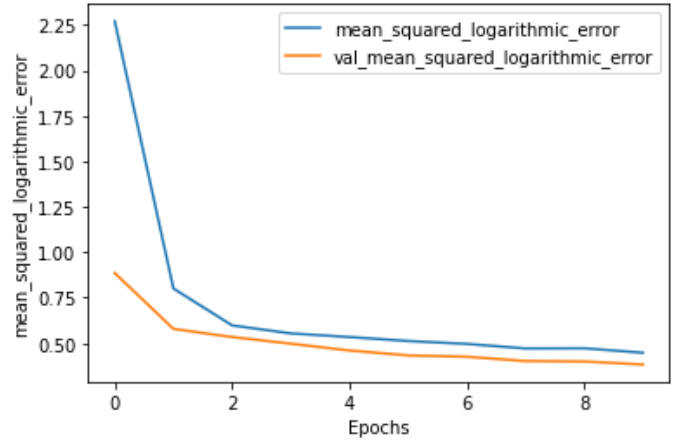
Sinir Ağları (Neural Network)

"Sinir ağları, karmaşık ve çok boyutlu verileri basitleştirir, değerler arasında kolay işlem yapılmasını sağlar." [4] Sinir ağlarında kullanılan model *Grafik NN1*'de belirtilmiştir.



Grafik NN1 [4]

İlgili örnek için 3 gizli katman, relu aktivasyonu ve 0.01 learning_rate ile eğitilen ve 10 epoktan oluşan yapay sinir ağı modelinde elde edilen sonuçlar ışığında elde edilen Epok sayısı ile Mean Squared Logarithmic Error değeri arasındaki ilişki *Grafik NN2*'de belirtilmiştir.



Grafik NN2

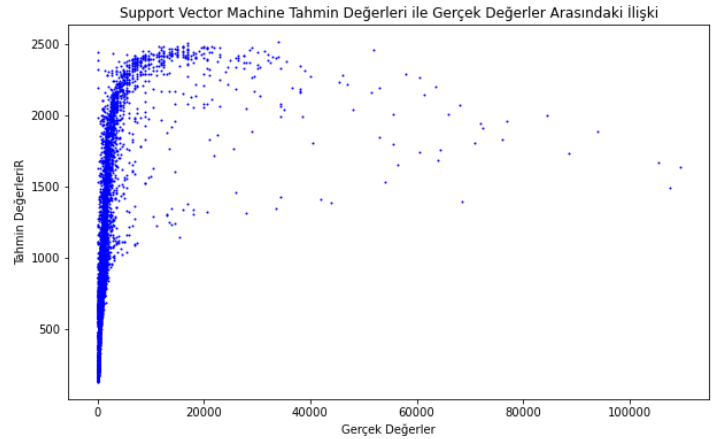
İlgili değerlerle eğitilen yapay sinir ağı modelinden elde edilen sonuçlar ise *Tablo NN1*'de belirtilmektedir.

Mean Absolute Error	928.4198
Mean Squared Error	17338952.2863
Root Mean Squared Error	4164.0067
R2 Score	0.6978

Tablo NN1

Destek Vektör Makineleri (Support Vector Machines)

"Destek vektör makineleri (DVM) (Cortes & Vapnik, 1995), tüm iyi bilinen veri madenciliği algoritmalarında en sağlam ve doğru yöntemlerdendir. Vapnik tarafından 1990'larda geliştirilen DVM'ler, istatistiksel öğrenme teorisini temel alan sağlam bir teorik temele dayanır, yalnızca bir düzine eğitim örnekleri gerektirir ve genellikle boyut sayısına duyarsızdır." [4] İlgili örnek Destek Vektör makinesi ile eğitildiğinde elde edilen grafik *Grafik SVM1*'deki gibidir.



Grafik SVM1

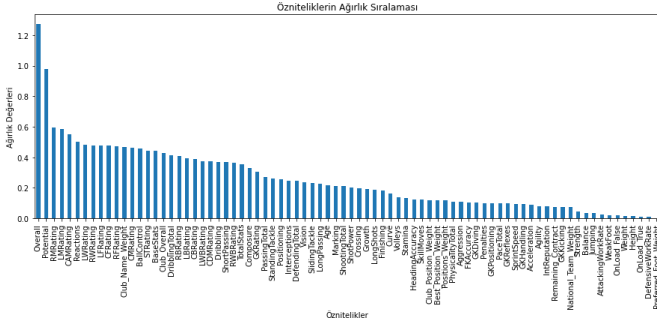
Destek Vektör Makineleri ile eğitilen ilgili veri setinin sonuçları ise *Tablo SVM1*'deki gibidir.

Train Score	0.0123
Test Score	0.0157
Mean Absolute Error	1917.6604
Mean Squared Error	52033226.8387
Root Mean Squared Error	7213.4060
R2 Score	0.0157

Tablo SVM1

C. Öznitelik Seçimi (Feature Selection)

Bir futbolcunun piyasa değerini belirleyen birçok etken vardır. Çok fazla etken seçilmesi ya da gereğinden az etken seçilmesi futbolcunun piyasa değerinin belirlenmesinde hatalı sonuçları doğurabilmektedir. Dolayısıyla etkenler doğru seçilmelidir. Belirli özniteliklerin ağırlıklarıyla beraber güncellenen veri setinden en uygun özniteliklerin seçilebilmesi amacıyla SelectPercentile fonksiyonu kullanılmıştır. Bu fonksiyon yardımıyla öznitelikler arasından sonuca en çok etkisi olanlardan en az olanlara doğru sıralama yapılmıştır. *Grafik O1*



Grafik O1

Bu grafik ışığında sonuca en çok etkisi bulunan özniteliklerden %20'si olmak üzere 'Overall', 'Potential', 'BaseStats', 'BallControl', 'Reactions', 'STRating', 'LWRating', 'LFRating', 'CFRating', 'RFRating', 'RWRating', 'CAMRating', 'LMRating', 'CMRating', 'RMRating', 'Club_Name_Weight' öznitelikleri bağımlı değişken (X) olarak seçilmiştir. 'ValueEUR' değeri ise bağımsız değişken (Y) olarak tahmin edilmeye çalışılmıştır. Bu özniteliklere bağlı olarak futbolcunun piyasa değerinin numerik değer olarak tahmin edilmesi amaçlanmıştır.

Not: İlgili veri setindeki bağımsız değişkenin (Y) değerleri çok yüksek olduğundan tüm değerler 0.001 ile çarpılarak kullanılmıştır.

III. SONUÇLAR VE TARTIŞMA

Bir futbolcunun alınıp satılabilmesi ve takımlara transfer olabilmesi için gerekli olan ücretin belirlenmesine futbolcunun piyasa değeri denilmektedir. Bu değer belirlenmesini de etkileyen birçok etken bulunmaktadır. İlgili veri setinde yer alan 91 öznitelikten isim, ID, fotoğraf, forma numarası gibi etkisi bulunmayan öznitelikler çıkarıldıktan sonra elde edilen 80 öznitelik için ağırlığı en yüksek olan öznitelikler belirlenmiştir. Bu özniteliklerle birlikte 5 farklı algoritma ile model oluşturulup piyasa değeri tahmini yapılmıştır.

Algoritma	R2 Skoru
KNN	0.8685
Doğrusal Regresyon	0.5254
Karar Ağacı	0.8975
Sinir Ağları	0.6978
Destek Vektör Makineleri	0.0157

Eğitilen modeller sonucunda elde edilen sonuçlar belirlenmiştir. Bu sonuçlar ışığında KNN algoritması ile Karar Ağacı modelinin birbirine yakın değerler çıkardığı ancak Karar Ağacı modelinin daha güvenilir sonuçlar ürettiği görülmektedir. İlgili modellerden sonra sırayı Sinir Ağları, Doğrusal Regresyon ve Destek Vektör Makinelerinin izlediği görülmektedir. Bununla birlikte Destek Vektör Makineleri algoritmasının ilgili veri seti için başarılı sayılabilecek bir sonuç üretmediği de görülmektedir. İlgili modellerle ilgili sıralama yapılacak olursa sıralamanın Karar Ağacı (Decision Tree) > KNN > Sinir Ağları (Neural Networks) > Doğrusal Regresyon (Linear Regression) > Destek Vektör Makineleri (Support Vector Machine) şeklinde olduğu görülmektedir.

KAYNAKLAR

- [1] K. Bang, D. Weissman, J. Wood, and M. Attfield, "Tuberculosis mortality by industry in the united states, 1990–1999," *The International Journal of Tuberculosis and Lung Disease*, vol. 9, no. 4, pp. 437–442, 2005.
- [2] Y. Özkan and V. M. Yöntemleri, "Papatya yayıncılık eğitim," *İstanbul, Mayıs-2008*, 2013.
- [3] M. Aydoğan and A. Karcı, "Meslek yüksekokulu öğrencilerinin başarı performanslarının makine öğrenmesi yöntemleri ile analizi," in *International Symposium on Multidisciplinary Studies and Innovative Technologies*, 2018.
- [4] E. NAMLI, Ü. Ramazan, and G. Ecem, "Fiyat tahminlemede makine öğrenmesi teknikleri ve doğrusal regresyon yöntemlerinin kıyaslanması; türkiye'de satılan ikinci el araç fiyatlarının tahminlenmesine yönelik bir vaka çalışması," *Konya Mühendislik Bilimleri Dergisi*, vol. 7, no. 4, pp. 806–821, 2019.
- [5] S. R. Safavian and D. Landgrebe, "A survey of decision tree classifier methodology," *IEEE transactions on systems, man, and cybernetics*, vol. 21, no. 3, pp. 660–674, 1991.
- [6] B. Charbuty and A. Abdulazeez, "Classification based on decision tree algorithm for machine learning," *Journal of Applied Science and Technology Trends*, vol. 2, no. 01, pp. 20–28, 2021.