AKDENİZ ÜNİVERSİTESİ
SOSYAL BİLİMLER ENSTİTÜSÜ

Tuba SARI

MAKİNE ÖĞRENMESİ YÖNTEMLERİ İLE NBA OYUNCU DEĞERİ ÖNCÜLERİNİN TAHMİN EDİLMESİ

Yönetim Bilişim Sistemleri Anabilim Dalı
Yüksek Lisans Tezi

Antalya, 2023

AKDENİZ ÜNİVERSİTESİ
SOSYAL BİLİMLER ENSTİTÜSÜ

Tuba SARI

MAKİNE ÖĞRENMESİ YÖNTEMLERİ İLE NBA OYUNCU DEĞERİ ÖNCÜLERİNİN TAHMİN EDİLMESİ

Danışman
Doç. Dr. Kerim Kürşat ÇEVİK

Yönetim Bilişim Sistemleri Anabilim Dalı
Yüksek Lisans Tezi

Antalya, 2023

T.C.

Akdeniz Üniversitesi

Sosyal Bilimler Enstitüsü Müdürlüğüne,

Tuba Sarı'nın bu çalışması, jürimiz tarafından Yönetim Bilişim Sistemleri Anabilim Dalı Tezli Yüksek Lisans Programı tezi olarak kabul edilmiştir.

Başkan	: Doç. Dr. Emre DANDIL	(İmza)
Üye (Danışman)	: Doç. Dr. Kerim Kürşat ÇEVİK	(İmza)
Üye	: Dr. Öğr. Üyesi Ahmet TÜRKMEN	(İmza)

Tez Başlığı:	Makine Öğrenmesi Yöntemleri ile NBA Oyuncu Değeri Öncülerinin Tahmin Edilmesi

Onay: Yukarıdaki imzaların, adı geçen öğretim üyelerine ait olduğunu onaylarım.

Tez Savunma Tarihi : 19/09/2023

Mezuniyet Tarihi :28/09/2023

AKADEMİK BEYAN

Yüksek Lisans Tezi olarak sunduğum “Makine Öğrenmesi Yöntemleri ile NBA Oyuncu Değeri Öncülerinin Tahmin Edilmesi” adlı bu çalışmanın, akademik kural ve etik değerlere uygun bir biçimde tarafımca yazıldığını, yararlandığım bütün eserlerin kaynakçada gösterildiğini ve çalışma içerisinde bu eserlere atıf yapıldığını belirtir; bunu şerefimle doğrularım.

Tuba SARI

T.C.
AKDENİZ ÜNİVERSİTESİ
SOSYAL BİLİMLER ENSTİTÜSÜ

25 /09/ 2023

TEZ ÇALIŞMASI ORİJİNALLİK RAPORU BEYAN BELGESİ

Öğrenci Bilgileri	
Adı-Soyadı	Tuba SARI
Öğrenci Numarası	202252096601
Anabilim Dalı	Yönetim Bilişim Sistemleri
Programı	Yönetim Bilişim Sistemleri Tezli Yüksek Lisans
Danışman Öğretim Üyesi Bilgileri	
Unvanı, Adı-Soyadı	Doç. Dr. Kerim Kürşat ÇEVİK
Yüksek Lisans Tez Başlığı	Makine Öğrenmesi Yöntemleri ile NBA Oyuncu Değeri Öncülerinin Tahmin Edilmesi
Turnitin Bilgileri	
Ödev Numarası	2176232035
Rapor Tarihi	11.09.2023
Benzerlik Oranı	Alıntılar hariç: % 4 Alıntılar dahil: %4

SOSYAL BİLİMLER ENSTİTÜSÜ MÜDÜRLÜĞÜNE,

Yukarıda bilgileri bulunan öğrenciye ait tez çalışmasının a) Kapak sayfası, b) Giriş, c) Ana Bölümler ve d) Sonuç kısımlarından oluşan toplam 108 sayfalık kısmına ilişkin olarak Turnitin adlı intihal tespit programından Sosyal Bilimler Enstitüsü Tez Çalışması Orijinallik Raporu Alınması ve Kullanılması Uygulama Esaslarında belirlenen filtrelemeler uygulanarak yukarıdaki detayları verilen ve ekte sunulan rapor alınmıştır.
Danışman tarafından uygun olan seçenek işaretlenmelidir:
(x) Benzerlik oranları belirlenen limitleri aşmıyor ise:
Yukarıda yer alan beyanın ve ekte sunulan Tez Çalışması Orijinallik Raporunun doğruluğunu onaylarım.
( ) Benzerlik oranları belirlenen limitleri aşıyor, ancak tez/dönem projesi danışmanı intihal yapılmadığı kanısında ise:
Yukarıda yer alan beyanın ve ekte sunulan Tez Çalışması Orijinallik Raporunun doğruluğunu onaylar ve Uygulama Esaslarında öngörülen yüzdelik sınırlarının aşılmasına karşın, aşağıda belirtilen gerekçe ile intihal yapılmadığı kanısında olduğumu beyan ederim.

Gerekçe:

Benzerlik taraması yukarıda verilen ölçütlere uygun olarak tarafımca yapılmıştır. İlgili tezin orijinallik raporunun uygun olduğunu beyan ederim.

Danışman Öğretim Üyesi
Doç. Dr., Kerim Kürşat ÇEVİK
İmza

İ Ç İ N D E K İ L E R

ŞEKİLLER LİSTESİ ..... **iii**
TABLOLAR LİSTESİ ..... **vi**
KISALTMALAR LİSTESİ ..... **vii**
ÖZET ..... **viii**
SUMMARY ..... **ix**
TEŞEKKÜR ..... **x**
GİRİŞ **1**

BİRİNCİ BÖLÜM
NBA

1.1. NBA Tarihi ve Yapısı 4
1.2. NBA Rekabet Dengesi 6
1.2.1. Salary Cap (Maaş Sınırı) 6
1.2.2. Lüks Vergilendirme 7
1.2.3. Draft Düzeni 7
1.2.4. Televizyon Gelirlerinin Paylaşımı 8
1.3. Basketbolda Oyuncu Pozisyonları 9
1.4. Win Share 10

İKİNCİ BÖLÜM
MAKİNE ÖĞRENMESİ

2.1. Makine Öğrenmesi Kavramı 12
2.2. Makine Öğrenmesi Kullanılan Alanlar 14
2.3. CRISP-DM (Veri Madenciliği için Endüstri Standart Süreci, Cross-Industry Standard Process for Data Mining) 17
2.4. Makine Öğrenmesinde Öğrenme Yöntemleri 20
2.4.1. Denetimli/Gözetimli Öğrenme (Supervised Learning) 20
2.4.2. Denetimsiz Öğrenme (Unsupervised Learning) 22
2.4.3. Pekiştirmeli/Takviyeli Öğrenme (Reinforcement Learning) 23
2.5. Makine Öğrenmesi Algoritmaları 24
2.5.1. Yapay Sinir Ağları (Neural Network) 24
2.5.2. Karar Ağacı (Decision Tree) 27
2.5.3. Karar Ağaçları Topluluk Öğrenmesi (Decision Tree Ensemble) 27
2.5.4. Rastgele Orman (Random Forest) 29
2.5.5. Lineer/ Doğrusal Regresyon (Linear Regression) 30
2.5.6. XGBoost Learning 31
2.6. Değerlendirme Ölçütleri 31
2.6.1. R Kare ve Düzeltilmiş R Kare (R-Squared ve Adjusted R-Squared) 32
2.6.2. Ortalama Kare Hatası (Mean Squared Error, MSE) ve Kök Ortalama Kare Hatası (Root Mean Squared Error, RMSE) 32
2.6.3. Ortalama Mutlak Hata (Mean Absolute Error, MAE) 33

ÜÇÜNCÜ BÖLÜM
NBA'DE MAKİNE ÖĞRENMESİ UYGULAMALARI İLE İLGİLİ YAPILAN ÇALIŞMALAR

3.1. NBA'de Maaş Tahminiyle İlgili Yapılan İstatiksel Çalışmalar 35
3.2. NBA'de Makine Öğrenmesi Uygulamaları ile İlgili Yapılan Çalışmalar 37
3.3. NBA'de Makine Öğrenmesi Uygulamaları Kullanılarak Maaş Tahminiyle İlgili Yapılan Çalışmalar 42

DÖRDÜNCÜ BÖLÜM
MATERYAL VE YÖNTEM

4.1. Araştırmanın Materyal ve Yöntemi 49
4.2. Veri Toplama Süreci 49
4.3. Veri Önişleme 51
4.4. Makine Öğrenmesi İçin Kullanılan Bilgisayar ve Program 52
4.5. Kullanılan CRISP-DM Yöntemi 53
4.6. Makine Öğrenmesi Uygulaması 54
4.6.1. Excel Reader Düğümü 54
4.6.2. Histogram Düğümü 56
4.6.3. Statistics Düğümü 57
4.6.4. Scatter Plot (Dağılım/Serpilme Grafiği) Düğümü 58
4.6.5. Color Manager Düğümü 59
4.6.6. Linear Correlation Düğümü 62
4.6.7. Column Filter 64
4.6.8. Missing Value Düğümü 64
4.6.9. Category to Number 65
4.6.10. Column Filter 66
4.6.11. Normalizer Düğümü 67
4.6.12. Number to String (PMML) 68
4.6.13. Partitioning 69
4.6.14. Lineer Regresyon 71
4.6.15. Yapay Sinir Ağları (MLP) 73
4.6.16. XgBoost 76
4.6.17. Rastgele Orman 77
4.6.18. Karar Ağaçları Topluluk Öğrenmesi Rastgele Orman 79
4.6.19. Karar Ağaçları Topluluk Öğrenmesi Gradient Boosting 81
4.6.20. Destek Vektör Makineleri (SVM) 84
4.6.21. Karar Ağacı 86
4.6.22. Model Doğrulama- Çapraz Doğrulama (Cross Validation) 88

BEŞİNCİ BÖLÜM
BULGULAR

5.1. Bulgular 96
SONUÇ 102
KAYNAKÇA 107
EK 1- NBA Takımları ve Kısaltmaları 120
EK 2- Veri Setindeki Değişkenler ve Anlamları 121
ÖZGEÇMİŞ 124

ŞEKİLLER LİSTESİ

Şekil 2.1 Makine Öğrenmesi Yapay Zekâ Arasındaki Bağlantı 12
Şekil 2.2 CRISP-DM Akış Şeması 17
Şekil 2.3 Makine Öğrenmesi Yöntemleri 20
Şekil 2.4 Yapay Sinir Hücresi Yapısı 25
Şekil 2.5 Karar Ağacı Yapısı 27
Şekil 2.6 Torbalama ve Güçlendirme Yöntemleri 28
Şekil 2.7 Karar Ağacı ve Rastgele Orman Algoritması Yapıları 29
Şekil 4. 1 Metodoloji Akış Şeması 49
Şekil 4.2 Excel Reader Ayarlar 54
Şekil 4.3 Veri Seti 55
Şekil 4.4 Excel Reader Düğümü Değişkenler 55
Şekil 4.5 Histogram Salary 21-22 56
Şekil 4.6 Histogram WS 21-22 57
Şekil 4.7 Statistics Düğümü 58
Şekil 4.8 Statistics Düğümü Top/Bottom 58
Şekil 4.9 Scatter Plot Düğümü 59
Şekil 4.10 Scatter Plot Maaş Oyuncu Grafiği 59
Şekil 4.11 Color Manager Düğümü 60
Şekil 4.12 Scatter Plot Maaş ve ID 60
Şekil 4.13 Scatter Plot Maaş ve WS 61
Şekil 4.14 Scatter Plot Maaş ve Deneyim 61
Şekil 4.15 Scatter Plot Maaş ve Pozisyon 62
Şekil 4.16 Scatter Plot Maaş ve PTS 62
Şekil 4.17 Lineer Correlation 63
Şekil 4.18 Column Filter Düğümü 64
Şekil 4.19 Missing Value Düğümü 65
Şekil 4.20 Category to Number Düğümü 66
Şekil 4.21 Column Filter Düğümü 66
Şekil 4.22 Normalizer Düğümü 67
Şekil 4.23 Normalize Veri Seti 68
Şekil 4.24 Number to String Düğümü 69
Şekil 4.25 Partitioning Düğümü 69

Şekil 4.26 Knime Programında Görünüm 70
Şekil 4.27 Modellerin Knime Programında Görünümü 71
Şekil 4.28 Lineer Regresyon Düğümü 72
Şekil 4.29 Lineer Regresyon Eğitim Sonuçları 73
Şekil 4.30 Lineer Regresyon Test Sonuçları 73
Şekil 4.31 MLP Learner Düğümü 74
Şekil 4.32 MLP Learner 74
Şekil 4.33 MLP Eğitim Sonuçları 75
Şekil 4.34 MLP Test Sonuçları 75
Şekil 4.35 XGBoost Düğümü 76
Şekil 4.36 XGBoost Düğümü Sonuçları Eğitim 77
Şekil 4.37 XGBoost Düğümü Sonuçları Test 77
Şekil 4.38 Rastgele Orman Düğümü 78
Şekil 4.39 Rastgele Orman Sonuçları Eğitim 79
Şekil 4.40 Rastgele Orman Sonuçları Test 79
Şekil 4.41 Topluluk Öğrenmesi Düğümü 80
Şekil 4.42 Topluluk Öğrenmesi Eğitim Sonuçları 81
Şekil 4. 43 Topluluk Öğrenmesi Test Sonuçları 81
Şekil 4.44 Gradient Boosted Trees Düğümü 82
Şekil 4.45 Gradient Boosted Trees Eğitim Sonuçları 83
Şekil 4.46 Gradient Boosted Trees Test Sonuçları 83
Şekil 4.47 SVM ve Karar Ağacı Modellerinin Knime Programında Görünümü 84
Şekil 4.48 SVM Learner Düğümü 84
Şekil 4.49 String to Number Düğümü 85
Şekil 4.50 SVM Eğitim Sonuçları 86
Şekil 4.51 SVM Test Sonuçları 86
Şekil 4.52 Karar Ağacı Düğümü 87
Şekil 4.53 Karar Ağacı Eğitim Sonuçları 88
Şekil 4.54 Karar Ağacı Test Sonuçları 88
Şekil 4. 55 Çapraz Doğrulama 89
Şekil 4.56 X-Partitioner 89
Şekil 4.57 X-Aggregator Düğümü 90
Şekil 4.58 Lineer Regresyon 90
Şekil 4.59 Lineer Regresyon Sonuç 91

Şekil 4.60 MLP Sonuçları 91
Şekil 4.61 XgBoost Learning Sonuçları 92
Şekil 4.62 Rastgele Orman Sonuçları 93
Şekil 4.63 Karar Ağaçları Topluluk Öğrenmesi Rastgele Orman Sonuçları 93
Şekil 4.64 Karar Ağaçları Topluluk Öğrenmesi Gradient Boosting Sonuçları 94
Şekil 4.65 SVM Sonuçları 95
Şekil 4.66 Karar Ağacı Sonuçları 95

TABLOLAR LİSTESİ

Tablo 1.1 2023 Yılı NBA Takım Listesi 5
Tablo 2.1 Biyolojik Sinir Sistemi Elemanları ve Yapay Sinir Sisteminde Karşılıkları 24
Tablo 3.1 Literatür taraması 44
Tablo 4.1 Playofflara Katılmayan Oyuncular ve Takımları 50
Tablo 5.1 Modellerin Değerlendirilmesi 98
Tablo 5.2 İyileştirme Sonrası Modellerin Değerlendirilmesi 100

KISALTMALAR LİSTESİ

CBA	Collective Bargaining Agreement
CRISP-DM	Cross Industry Standard Process for Data Mining
DEF	Defansif Ribaund (Defensive Rebound Differential Per Game)
FG	2 Sayılık Atış (Field Goal Percentage Differential Per Game)
FT	Serbest Atış (Free Throw Percentage Differential Per Game)
K-NN	K-Nearest Neighbors (K-En Yakın Komşu Algoritması)
LASSO	Least Absolute Shrinkage and Selection Operator
MAE	Mean Absolute Error (Ortalama Mutlak Hata)
MLB	Major League Baseball
MLP	Multilayer Perceptron (Çok Katmanlı Algılayıcılar)
MSE	Mean Squared Error (Ortalama Kare Hatası)
NBA	National Basketball Association
NFL	Nationa Football League
NHL	National Hockey League
NLP	Natural Language Processing (Doğal Dil İşleme)
P	3 Sayılık Atış (3-Point Field Goal Percentage Differential Per Game)
PCA	Temel Bileşenler Analizi (The Principal Component Analysis)
PTS	Total Points
PER	Oyuncu Verimlilik Derecesi (Player Efficiency Rating)
RMSE	Root Mean Squared Error (Kök Ortalama Kare Hatası)
SVM	Support Vector Machines (Destek Vektör Makineleri)
TO	Kontra Atak (Turnover Differential Per Game)
VORP	Yedek Oyuncuya Göre Üstünlük Değeri (Value Over Replacement Player)
YSA	Yapay Sinir Ağları

ÖZET

Spor tüm dünya tarafından takip edilen bir endüstri olması sebebiyle birçok bilim dalı tarafından araştırılmıştır. Spor gibi çok değişkenli ve büyük verilerin olduğu bir alanı analiz etmeyi hedefleyen akademik çalışmalar da oldukça fazladır. Makine öğrenmesi yöntemlerini spor verileri üzerinde kullanarak yapılan araştırmalar literatürde sıklıkla gözlemlenmektedir. Toplumun en çok takip ettiği spor dallarından biri basketboldur. National Basketball Association (NBA), Amerika Birleşik Devletleri merkezli profesyonel basketbol ligi olarak bilinmektedir. NBA, 1946 yılında kurulmuş olup, günümüzde dünyada en çok takip edilen basketbol organizasyonu olarak kabul edilmektedir. En çok takip edilen basketbol ligi olmasından dolayı araştırmada NBA verilerinden yararlanılmıştır.

Bu tez çalışmasında oyuncuların istatistiklerinin maaşlarına etkisi, oyuncuların Win Share (kazanma payı) değerleri makine öğrenimi yöntemleri ile analiz edilmiştir. Tez çalışmasında oyuncuların sahadaki performansları, takımların bordroları ve maaşa etki edecek diğer değişkenler ikincil verilerden toplanmıştır. Win Share değişkenini sekiz farklı makine öğrenmesi algoritması kullanarak tahmin yapılmıştır. Makine öğrenmesi modellerinden Lineer Regresyon, Yapay Sinir Ağları, XgBoost Learning, Rastgele Orman, Karar Ağaçları Topluluk Öğrenmesi Rastgele Orman ve Gradient Boosting, SVM ve Karar Ağaçları olmak üzere 8 farklı algoritma kullanılarak modelleme yapılmış R^2, MAE, MSE, RMSE ve MAPE hata ölçüm sonuçları karşılaştırılarak analiz edilmiştir. İlk analizden sonra modelde Çapraz Doğrulama yapılmış sonuçlar tekrar hesaplanmıştır. Bu algoritmalar içinden Win Share'i en başarılı şekilde tahmin eden algoritmalar tespit edilmiştir

Bu tez çalışmasında çapraz doğrulama sonrası model başarısı en yüksek modeller R^2'ye göre 0,839 ile Rastgele Orman, 0,824 ile Karar Ağaçları Topluluk Öğrenmesi Gradient Boosting, 0,819 ile Karar Ağaçları Topluluk Öğrenmesi Rastgele Orman olarak tespit edilmiştir. Bu çalışmada hesaplanan Win Share kullanılarak gerçek bir değer elde edilebilecektir. Başarı oranı yüksek tahminleme yapan algoritma kullanılarak bir maaş hesabına ulaşılabilecektir. Bu şekilde oyuncuların hak ettiği maaşa yakın gerçek bir bulgu elde edilebilecektir.

Anahtar Kelimeler: Makine Öğrenmesi, Yapay Zekâ, Tahminleme, Veri Madenciliği, NBA

SUMMARY

PREDICTING NBA PLAYER VALUE PRE-INDICATORS USING MACHINE LEARNING METHODS

Since sports is an industry followed all over the world, it has been researched by many branches of science. There are also many academic studies that aim to analyze a field with multiple variables and large data, such as sports. Research using machine learning methods on sports data is frequently observed in the literature. One of the sports most followed by society is basketball. National Basketball Association (NBA) is known as the professional basketball league based in the United States. NBA was founded in 1946 and is now considered the most followed basketball organization in the world. NBA data was used in the research because it is the most followed basketball league.

In this thesis study, the effect of players' statistics on their salaries and the players' Win Share values were analyzed using machine learning methods. In the thesis, the performance of the players on the field, the payrolls of the teams and other variables that will affect the salary were collected from secondary data. The Win Share variable was predicted using eight different machine learning algorithms. R2, MAE, MSE, RMSE and MAPE error measurement results were modeled using 8 different algorithms, including Linear Regression, Artificial Neural Networks, XgBoost Learning, Random Forest, Decision Trees, Ensemble Learning, Random Forest and Gradient Boosting, SVM and Decision Trees, among machine learning models. were compared and analyzed. After the first analysis, the model was cross-validated and the results were recalculated. Among these algorithms, the ones that predict Win Share most successfully have been identified.

In this thesis, the models with the highest model success after cross-validation were determined as Random Forest with 0.839 according to R2, Decision Trees Ensemble Learning Gradient Boosting with 0.824, and Decision Trees Ensemble Learning Random Forest with 0.819. A real value can be obtained using the Win Share calculated in this study. A salary calculation can be achieved by using an algorithm that makes predictions with a high success rate. In this way, a real finding close to the salary that the players deserve can be obtained.

Keywords: Machine Learning, Artificial Intelligence, Estimation, Data Mining, NBA

TEŞEKKÜR

Tez çalışması boyunca değerli yardım ve yönlendirmeleriyle, desteğini esirgemeyen tez danışmanım Doç. Dr. Kerim Kürşat ÇEVİK'e teşekkür ederim.

Lisans ve yüksek lisans eğitimim boyunca bilgileriyle ışık tutan Akdeniz Üniversitesi Yönetim Bilişim Sistemleri Bölümü'ndeki tüm hocalarıma her zaman gösterdikleri destek ve inanç için teşekkür ederim.

Mesafelerde bile varlıklarını hep yanımda hissettiğim, inancımı veya kararlılığımı asla kaybetmememi sağlayan manevi desteği sonsuz olan dostum Şeyda KİĞILI ve arkadaşlarıma teşekkür ederim.

Haklarını ödeyemeyeceğim annem, babam ve kardeşime yıllar süren benzersiz maddi ve manevi destek ve rehberlikleri için ayrıca akademik hayatımda yön bulmamda sağladıkları olanaklarla birlikte bu yolda bana her zaman inandıkları için teşekkür ederim.

GİRİŞ

Spor son yıllarda tüm dünya tarafından takip edilmektedir. Güçlü bir endüstriye dönüşen spor büyük bir finans kaynağı haline gelmiştir. Bu finans kaynağını daha iyi analiz etmek önem arz etmeye başlamıştır. Spor bilimleri, birçok farklı bilim dalı tarafından araştırılmış ve istatistiksel yöntemler kullanılarak analizler yapılmıştır. Ancak, teknolojinin gelişmesiyle birlikte veri bilimi gibi yeni yöntemler de ortaya çıkmıştır ve bu yöntemlerle analizler daha da güçlendirilerek gelecek tahminleri yapılmaya başlanmıştır. Özellikle futbol ve basketbol gibi popüler spor dallarında, veri bilimi ve makine öğrenmesi teknikleri kullanarak yapılan analizler büyük avantajlar sağlamaktadır. Bu alanda yapılan tahminleme çalışmaları henüz sınırlı olsa da bu alandaki gelişmeler ve yenilikler sayesinde spor bilimleri araştırmalarında daha da ileriye gidileceği düşünülmektedir.

Toplumun en çok takip ettiği spor dallarından biri basketboldur. Dünyada en çok ilgi gören ve bilinen basketbol ligi National Basketball Association (NBA)'dir. NBA, Amerika Birleşik Devletleri merkezli profesyonel basketbol ligi olarak bilinmektedir. NBA'de 30 takım bulunmaktadır ve NBA Kuzey Amerika'daki dört ana spor liginden biridir (Hentilä, 2019: 7). Her yıl düzenlenen Playoff maçları ve NBA Finalleri ile sonuçlanan bir sezona sahiptir. Ayrıca NBA All-Star maçı gibi özel etkinlikler de düzenlenmektedir. NBA, yıldız basketbolcuların yetiştiği birçok ülkede hayran kitlesine sahip ve dünya çapında milyarlarca dolarlık bir endüstridir.

NBA'de çeşitli metriklerle farklı konularda analizler yapılmaktadır. Oyuncular, koçlar, takım sahipleri ve analistler bu analizler sayesinde takımı ve oyuncuları daha iyi anlayabilmektedirler. Oyunlar kadar sporcuların performansları da ölçülmeye çalışılmaktadır. Oyunları ve oyuncuları iyi analiz ederek doğru takımı yaratmak, doğru zamanda doğru oyuncuyu oyuna sokarak optimum kararları geliştirebilmek, takımların ve oyuncuların zayıf ve güçlü yönlerini belirleyerek optimum antrenman programları oluşturmak, oyuncuların maaşlarını tahminlemek gibi birçok konu veri analizleri ve makine öğrenmesi yöntemleriyle yapılabilmektedir.

NBA'de sadece bir sezonda bir takım 82 maç oynamaktadır. Bu oyunlarda büyük miktarda veriler oluşmaktadır. NBA verileri ligin kuruluşu olan 1946 yılından beri tutulmaktadır (NBA Media Ventures, 2023b). Veri toplama ve saklama yöntemleri zamanla değişti ve gelişti. İstatistiklerin daha ayrıntılı hale getirilmesi ve dijital veri tabanlarının kullanılmaya başlanmasıyla birlikte, NBA istatistikleri daha kapsamlı hale gelmiştir. Oyunların

ve oyuncuların da verilerini tutan NBA kendi resmî sitesinde şeffaf bir şekilde bu verileri yayınlamaktadır.

Spor bilimlerinde kullanılan veri bilimi sadece NBA tarafından değil, üniversiteler tarafından da araştırılmaktadır ve veri bilimiyle ilgili alanlarda internette yer alan birçok eğitici bloglarda, web sitelerinde kullanılmaktadır. Indiana Üniversitesi Luddy School of Informatics, Computing and Engineering ve Tübingen Üniversitesi gibi üniversitelerin Spor Analizi İçin Veri Bilimi gibi proje programları ve laboratuvarları olduğu gibi son zamanlarda üniversitelerin bu alanda dersler sunduğu görülmektedir.

NBA veri bilimine oldukça önem vermektedir. Takımların neredeyse hepsinde analiz departmanları bulunmaktadır. Veri bilimindeki gelişen araştırmalar sayesinde geçmiş yıllara göre çok daha etkili analizler yapılmaktadır (Janiesch vd., 2021: 685). NBA'deki veri analitiği kullanımı sayesinde takımlar artık kadrolarını buradan çıkan analizlere göre oluşturuyor ve oyuncuların yönetimini optimum şekilde kurgulayabilmektedirler.

NBA'de oyuncuların maaş tahmini, geniş bir konu yelpazesine ve birçok disipline uygulanabilir. Bu tahminler, ekonomi, finans, işletme, spor yönetimi, spor hukuku ve sosyoloji gibi alanlarda derinlemesine araştırmalara ve analizlere yol açabilir. NBA büyük bir endüstri olduğu için ekonomi ve finans alanında birçok araştırma konusu sunar. Oyuncuların maaş tahminleri, spor endüstrisindeki gelir dağılımını, oyuncuların değerini belirleyen faktörleri ve maaş yapısının nasıl şekillendiğini anlamak için kullanılabilir.

Maaş tahminleri, takımların maaş bütçelerini yönetme, maaş pazarlığı stratejilerini geliştirme ve oyuncu tercihlerini belirleme süreçlerine ışık tutar. Ayrıca maaş tahminleri, takımların kaynaklarını etkili bir şekilde kullanmalarını sağlayarak sürdürülebilir bir başarı elde etmelerine yardımcı olabilir.

NBA'de oyuncuların maaş tahminleri, spor hukuku ve düzenlemeleriyle ilgilenen akademik çalışmalarda da yer alır. Maaş tahminleri oyuncuların maaşlarını belirleme sürecindeki adalet ve dengeyi, maaş sınırlamalarının etkisini ve sözleşme düzenlemelerini değerlendirmede kullanılabilir. Ayrıca maaş tahminleri, kolektif pazarlık anlaşmaları gibi konularla ilişkili olarak spor endüstrisindeki maaş eşitsizliklerini araştırmak için de önemli bir kaynak olabilir.

Bu tez çalışmasında NBA'de oyuncuların istatistiklerinin maaşlarına etkisi, oyuncuların Win Share (kazanma payı) değerleri makine öğrenimi yöntemlerinin etkisini araştırmayı amaçlamaktadır. Tezin hipotezine göre, makine öğrenmesi yöntemleri, NBA oyuncularının maaş tahmini yaparken istatistiksel verileri, performans ölçütlerini ve sosyal medya etkileşimleri gibi çok çeşitli faktörleri dikkate alarak geleneksel tahmin yöntemlerine kıyasla

daha doğru ve hassas tahminler yapabilir. Önerilen tezde farklı makine öğrenmesi algoritmaları kullanılarak maaş tahmini modelleri oluşturulacak ve bu modellerin performansı değerlendirilecektir.

Bu tez çalışması, NBA oyuncularının takıma kazandırmasını ve maaşlarına etki eden faktörlerin belirlenmesi ve makine öğrenmesi algoritmalarının kullanılmasıyla bu faktörlerin analiz edilmesi konularında değerli bir katkı sunmayı amaçlamaktadır.

Bu çalışmanın sonuçları, NBA ve diğer spor ligleri için maaş yönetimi stratejilerini iyileştirme ve oyuncu değerlendirmelerini daha bilinçli bir şekilde yapma konusunda değerli bilgiler sağlayabilir. Ayrıca makine öğrenmesi ve spor analitiği alanındaki literatüre katkıda bulunarak, bu tekniklerin spor endüstrisindeki uygulamalarını daha da geliştirebilir

Yapılan bu tez çalışmasında NBA'de oyuncu performanslarının çeşitli (Yapay Sinir Ağları, Karar Ağacı, Karar Ağaçları Topluluk Öğrenmesi, Rastgele Orman, Lineer Regresyon, XGBoost Learning) makine öğrenmesi yöntemleri kullanılarak maaşlara, Win Share (kazanma payı) değerleri ve oyuncuların all star (yıldız oyuncu) olma durumlarına olan etkisi araştırılmıştır.

Çalışmanın birinci bölümünde NBA tarihi, yapısı genel hatlarıyla yer almaktadır.

İkinci bölümde makine öğrenmesi kavramı, makine öğrenmesinin kullanıldığı alanlar, veri madenciliği metodolojilerinden kullanılan CRISP-DM yöntemi anlatılacaktır. İkinci bölümün devamında makine öğrenmesi adımları ve algoritmalarına yer verilmiştir.

Üçüncü bölümde NBA'deki verilerle makine öğrenmesi uygulamaları ile ilgili yapılan tez ve diğer çalışmalara değinilmiştir.

Dördüncü bölüm kapsamında makine öğrenmesinin NBA verileri üzerine uygulanması işlenmiştir. Verilerin toplanması, önişleme adımları, Knime programıyla uygulanan algoritmalar da olmak üzere yöntem ve materyaller açıklanmıştır.

Bu aşamadan sonra elde edilen modeli en iyi açıklayan değişkenler, maaşa olan etkileri, oyuncuların takıma kazandırma payları ile ilgili elde edilen bulgulara yer verilmiştir. Birden fazla makine öğrenmesi tekniği kullanılarak elde edilen sonuçlar diğer algoritmalardan elde edilen sonuçlar ile karşılaştırılmıştır. Bütün bulgular incelenerek optimum değeri bulmak amacıyla en uygun modeller ve değişkenler ortaya konulmuştur.

BİRİNCİ BÖLÜM
NBA

1.1. NBA Tarihi ve Yapısı

Amerika Ulusal Beyzbol Ligi (Major League Baseball, MLB), Ulusal Basketbol Birliği (National Basketball Association, NBA), Ulusal Futbol Ligi (National Football League, NFL) ve Ulusal Hokey Ligi (National Hockey League, NHL) Kuzey Amerika'daki dört ana spor ligidir. NBA, MLB, NFL ve NHL genellikle "Dört Büyük" olarak anılmaktadır. NBA dünyanın her yerinden oyuncuları bünyesinde bulundurmaktadır ve basketbol hızla Kuzey Amerika'daki en popüler spor haline gelmiştir. Her NBA takımının yurt dışından oyuncular içermesi nedeniyle lig, Kuzey Amerika dışında da oldukça büyük bir hayran kitlesini kendine çekmektedir. NBA takımları, ligin artan popülaritesi nedeniyle her yıl büyük miktarda para kazanmaktadır. Bu nedenle, NBA şu anda en iyi dört profesyonel spor ligi arasında en yüksek sözleşmelere sahip konumdadır (Hentilä, 2019: 7).

1946'da kurulan NBA 11 takımla kurulmuştur (NBA Media Ventures, 2023b). İlk başta Amerika Basketbol Birliği (Basketball Association of America) ismiyle kurulup daha sonra Ulusal Basketbol Ligi (National Basketball League, NBL) ile birleşmesiyle şu anki ismi olan NBA adını almıştır. 1949 da NBL'den 6 takımın daha katılmasıyla 17 takımla devam etmiştir. Yeni lig hızla takım kaybetmiş ve 6 yıl sonra takım sayısı 8'e düşmüştür. 1966'dan 2004'e kadar lige toplam 21 takımın katılmasıyla tekrar canlanmıştır (NBA Hoops Online, 2023).

NBA' de şu anda 30 takım bulunmaktadır. Bu takımlar konferans olarak ayrılmıştır. Coğrafi konumlarından dolayı doğu ve batı konferansı olarak adlandırılmaktadır. Her bir grupta 5 olmak üzere toplam bir konferansta 15 takım bulunmaktadır. Doğu konferansı kendi içerisinde Atlantic (Atlantik), Central (Merkez) ve Southeast (Güneydoğu) olmak üzere üç gruba ayrılmaktadır. Batı konferansı ise Northwest (Kuzeybatı), Pacific (Pasifik), Southwest (Güneybatı) olarak ayrılmaktadır (NBA Media Ventures, 2023a). NBA'de yer alan tüm takımların konferans ve grup dağılımları Tablo 1.1 'de yer almaktadır. NBA takımları ve kısaltmaları EK 1'de yer almaktadır. NBA, Franchise sistemine göre yapılandırılmıştır. NBA'de yükselme ve küme düşme gibi sistemler bulunmaktadır. Yeni takımların Şampiyonluk Komisyonu tarafından onaylanmasını da gerektiren bu sistemler, Avrupa liglerine göre farklılık göstermektedir. NBA zaman içerisinde birçok takımı bünyesine dahil etmiş ya da çıkarmış ve şehirler arasında takımları değiştirmiştir (Del Corral vd., 2016: 82).

Tablo 1.1 2023 Yılı NBA Takım Listesi

Doğu Konferansı			Batı Konferansı		
Atlantic	Central	Southeast	Northwest	Pacific	Southwest
Boston Celtics	Chicago Bulls	Atlanta Hawks	Denver Nuggets	Golden State Warriors	Dallas Mavericks
Brooklyn Nets	Cleveland Cavaliers	Charlotte Hornets	Minnesota Timberwolves	LA Clippers	Houston Rockets
New York Knicks	Detroit Pistons	Miami Heat	Oklahoma City Thunder	Los Angeles Lakers	Memphis Grizzlies
Philadelphia 76ers	Indiana Pacers	Orlando Magic	Portland Trail Blazers	Phoenix Suns	New Orleans Pelicans
Toronto Raptors	Milwaukee Bucks	Washington Wizards	Utah Jazz	Sacramento Kings	San Antonio Spurs

Kaynak: NBA Media Ventures, 2023a

NBA' de bir sezonda bir takım 82 maç yapmaktadır. Her bir takım önce kendi grubu içerisinde 4 maç oynamaktadır. Üç ya da dört maçı kendi konferansından olan ama diğer gruplarla oynamaktadır. Bu maçların kimlerle olacağı loto usulü bir kurayla belirlenmektedir. Takım iki oyunu da diğer konferanstan takımlara karşı oynamaktadır. Sezon maçları nisan ayında sona ermektedir.

NBA Playoff'ları sezon bittikten sonra nisan ayında başlar. Her bir konferansın en çok maç kazanan 8 takımı playoff oyunlarına katılır. Playoff' lar normal sezon bittikten sonra başlar ve doğu ile batı konferansları arasında gerçekleşmektedir. Toplamda her bir konferanstan 8 olmak üzere toplam 16 takım yarışmaktadır. Kazanan takımın konferansı playoff galibi olur.

Playoff' lar 7 maç üzerinden oynanmaktadır. İki takım art arda oynar ve 4 galibiyete ulaşan o seriyi kazanır. 4 galibiyeti bulamayan takım elenir. Galip takım ise bir sonraki takımla eşleşir. Bu şekilde takımlar arasında turlar konferans finaline kadar devam etmektedir. Konferans finalinden sonra o konferansın şampiyonları belli olmaktadır (NBA Media Ventures, 2023c).

NBA Finali haziran ayında oynanır. Playoff' larda iki konferanstan şampiyon olan iki takım arasında gerçekleşir. Playoff' lardaki sistemin aynısı finallerde de bulunmaktadır ve 7 maç üzerinden oynanmaktadır. Dört galibiyeti olan takım NBA finalinin galibi olmaktadır. Kazanan takıma Şampiyonluk Yüzüğü verilmektedir (Kaimakamis, 2021: 17).

1.2. NBA Rekabet Dengesi

NBA rekabet dengesini sağlamaya oldukça önem vermektedir. Yıllar içinde komisyon, oyuncular ve franchiselar arasında tartışılarak belirlenen bazı kurallar bulunmaktadır. NBA'deki franchise sistemi, lige katılan her takımın bağımsız bir işletme olarak yönetilmesi anlamına gelir (de Saá Guerra vd., 2011: 20). Her takım bir "franchise" veya "lisans" olarak bilinir ve bu lisans takımın ligin bir parçası olarak faaliyet göstermesine ve oyunlara katılmasına izin verir. Televizyon gelirlerinin paylaşımı, draft düzeni, maaş sınırı (Salary Cap) ve lüks vergilendirme oyuncular ve franchiseların koyduğu kuralların en önemlileridir (Del Corral vd., 2016: 82).

1.2.1. Salary Cap (Maaş Sınırı)

NBA'de oyuncuların maaş sınırları diğer bir deyişle maaş tavanları (Salary Cap) bulunmaktadır. Takımlar, bu maaşları oyuncuya göre farklı oranlarda dağıtır. Bazı oyuncular takımın toplam maaş bordrosunun yüzde 20 ila 25'ini almaktadır. Daha fazla performansı olan oyuncular daha büyük sözleşmeler yapmaktadırlar. Bir oyuncunun kulüp için değeri, ne kadar iyi basketbol oynadığının ötesine geçebilir (Hentilä, 2019: 7).

Maaş sınırı uygulaması ilk olarak 1983'te uygulamaya konulmuştur. NBA'deki her takımın yıllık toplam ücret harcamaları için bir üst sınır getirilmesi uygulaması yapılmıştır (Gök vd., 2021: 1108). NBA'de maaş tavanı 2023-2024 sezonu için yönetim tarafından 134 milyon USD, vergi seviyesi 162 milyon USD olarak belirlenmiştir (NBA Media Ventures, 2023d).

Maaş sınırlaması uygulamasıyla takımlar arasında uçurum maaş farkı bulunmamaktadır ve kadrolarına alabileceği yıldız oyuncu sayısı arasında uçurum oluşmamaktadır. Bu sayede ligin rekabet dengesi korunmaktadır.

Maaş sınırlamasında bazı esneklikler bulunmaktadır. Salary Cap; Hard Cap ve Soft Cap olarak iki ana kategoriye ayrılmaktadır. Hard Cap, belirlenmiş bir sınırı aşmayı kesinlikle yasaklayan bir kavramdır. Soft Cap ise toplu iş sözleşmesinde yer alan istisnai durumlar ve lüks vergisi gibi uygulamalar sayesinde sınırın aşılmasına izin verilebilen bir kavramdır (Ötkan ve Çolakoğlu, 2017: 965).

NBA, oyuncuların istihdam koşullarını ve şartlarını belirlemek, kulüplerin, oyuncu birliğinin ve NBA'nin hak ve yükümlülüklerini düzenlemek için toplu iş sözleşmesi (Collective Bargaining Agreement, CBA) kullanmaktadır (Gök vd., 2021: 1108). Toplu iş sözleşmesi içeriğinde, lüks vergisi yanı sıra "Salary Cap Exceptions" başlığı altında belirtilen istisnalar vardır. Yer alan birçok farklı istisnai durum Soft Cap kavramını etkilemektedir (Ötkan ve Çolakoğlu, 2017: 967).

1.2.2. Lüks Vergilendirme

NBA takımlarının fazla harcamalarının bir sonucu olarak, takımları maaş bütçesi konusunda sınırlayan bir uygulama olarak lüks vergisi bulunmaktadır. Lüks vergisi maaş sınırının bir parçasıdır. NBA tarafından belirlenen o sezonun lüks vergisi maaş sınırının üstünde yer almaktadır. Lüks vergi eşiğine ulaşıp aşan herhangi bir takım, eşiğe ulaştıktan sonra harcadıkları her dolar için belirli bir miktar ödemek zorundadır (Goldrosen, 2018: 16).

Lüks vergisi cezasının uygulandığı üst sınır, maaş istisnaları bu sınıra dahil edildiğinden, gerçek maaş üst sınırından daha yüksektir (K. Sigler ve Compton, 2018: 3).

NBA'de 2023-2024 sezonu için yönetim tarafından vergi seviyesi 162 milyon USD olarak belirlenmiştir (NBA Media Ventures, 2023d).

Lüks vergi eşiğini aşmanın getirdiği yüksek cezalar, takımlarına yıldız oyuncuları doldurmaktan caydırmasına rağmen yine de ekiplerin kendi istekleri doğrultusunda harcama yapmasına izin vermektedir (Goldrosen, 2018: 13).

Lüks vergisiyle ilgili bazı istisnalar da bulunmaktadır. Her takım çaylaklarından biri için beş yıllık bir uzatma teklif edebileceği tasarlanmış oyuncu eklentisi yapabilmektedirler. Takımlar ayrıca çaylaklarla yapacakları sözleşmelerde minimum ölçeklerle imzalayabilirler. Takımların yapacağı çaylaklarla olan bu sözleşmelerde de lüks vergilendirme sınırı aşabilmelerine olanak sağlayan esneklikler bulunmaktadır. Başka bir istisna ise cap room (sermaye) olmadan lüks vergisi sınırının altında olan bir takımın bir oyuncuya 8.406 milyon USD'e kadar ödeme yapmasına izin vermektedir. Başka bir istisna ismini bu istisnanın uygulandığı ilk oyuncu olan Larry Bird'den alan Bird Rights'dır. Bird Rights ile takımların kadrolarında bulunan oyuncular sınırı aşsalar bile yeniden sözleşme imzalamasına olarak tanımaktadır. Bu istisnaya uygun oyuncular maksimum maaşa kadar herhangi bir miktarda sözleşme yapabilmektedirler. Bir oyuncunun alabileceği maksimum maaş miktarı, NBA'de oynadığı yıl sayısına ve maaş sınırına bağlıdır (K. Sigler ve Compton, 2018: 3).

NBA'de ayrıca asgari maaş istisnası, maksimum maaş için oyuncunun limiti aşsa bile önceki sözleşmesinin yüzde yüz beşi kadar maaşa sözleşme imzalayabilmesi gibi farklı istisnalar da bulunmaktadır (K. Sigler ve Compton, 2018: 4).

1.2.3. Draft Düzeni

NBA de draft, genellikle "NBA Draft" veya "NBA Entry Draft" olarak adlandırılan, National Basketball Association (NBA) takımlarının yeni oyuncular kazanmak için yaptığı bir etkinliktir. Genellikle mayıs ya da haziran ayında gerçekleştirilir ve her takım, diğer takımlardan birer sıra ile oyuncu seçebilmektedir. Seçim sırası, takımların ligdeki son

durumlarına göre belirlenir. NBA draft' ında seçilen oyuncular, genellikle kolej veya yurtdışından gelirler ve seçildikleri takıma katılarak yeni profesyonel bir kariyere başlamaktadırlar. NBA Playoff 'larını kaçıran 14 takım Draft Loto'ya katılmaya hak kazanır (NBA Media Ventures, 2023c).

NBA draft kurası, 14 takımın katıldığı, her takımın bir topunun lotoda yer aldığı bir çekiliş yoluyla gerçekleştirilir ve her takımın belirli bir şansı vardır. Bu topun numarası, seçilecek olan takımın draft sırasını belirler. Örneğin, ilk çıkan top birinci sıradaki takımın draft sırasını belirler, ikinci çıkan top ikinci sıradaki takımın draft sırasını belirler ve bu şekilde devam etmektedir.

Lotoda, her takım için birer toplamda 14 adet şans kartı kullanılır. Daha sonra, 1 ile 1.000 arasında rastgele bir sayı üreten ve bu sayının hangi şans kartıyla eşleştiğini belirleyen bir çekiliş yapılır.

Eşleştirme sonucunda, en az şans kartına sahip takım, en yüksek şansı kazanır ve ilk turda ilk seçimi yapma hakkını elde eder. Diğer takımlar da sırayla seçim yapma hakkı kazanırlar. Bu şekilde, genç yeteneklerin takımlara dağılımı yapılır. Ancak, son yıllarda NBA'de draft sistemi biraz değişmiştir ve şans kartları arasındaki şans oranları daha dengeli hale getirilmiştir. Bu draft sistemiyle, başarısız olan takımların gelecek sezonlar için en iyi yetenekleri seçme şansı artar ve böylece ligde biraz denge sağlanmaya çalışılır. Güncel olarak 2019'da belirlenen Draft sistemi uygulanmaktadır (NBA Media Ventures, 2023c). Draft, genç oyuncuların lige girmesinin birincil yoludur. 2018 itibariyle, oyuncuların NBA draftına hak kazanabilmesi için, oyuncunun en az 19 yaşında olması ve drafttan 60 gün önce uygunluğunu beyan etmesi gerekmektedir (Chun vd., 2020: 8).

ABD'deki üniversitelerden mezun olan drafta katılmak isteyen bütün oyuncular için farklı bir durum bulunmaktadır. Dört senelik okullardan mezun bu oyuncular direkt NBA drafta seçilme hakkı kazanmaktadırlar. Ancak dört yıllık üniversitelerden mezun olmayan herhangi bir oyuncu, profesyonel olmak istediğini belirttikten sonra NBA 'e başvurarak seçmelere katılma hakkını elde edebilir (Kıratlı, 2020: 13).

NBA draft, her yıl geleceğin yıldızlarına ve NBA takımlarının gelecekte ne tür sporculara yatırım yaptığına dair önemli bir gösterge olarak kabul edilmektedir.

1.2.4. Televizyon Gelirlerinin Paylaşımı

Televizyon yayıncılığından elde edilenler ünlü ve takip edilen spor takımları, ligler ve spor federasyonlarının asıl gelir kaynağını oluşturmaktadır (Gedik, 2022: 176). NBA'de televizyon yayınlarından elde edilen gelirler, ligin en önemli gelir kalemlerinden biridir. Bu

gelirlerin kaynakları arasında ulusal yayıncılar, yerel yayıncılar ve dijital yayınlar yer almaktadır. 2019 yılında, NBA'in ESPN ve TNT kanalları ile olan 9 yıllık anlaşması 24 milyar dolar değerindedir (Sandomir, 2014).

NBA yönetimi, maç yayını gelirlerini artırmak ve ligin marka değerini yükselterek tanıtımını geliştirmek amacıyla çeşitli çalışmalar yürütmektedir. Bu çalışmalar arasında NBA maçlarının ABD ve Kanada dışında da denizaşırı ülkelerde televizyonda oynatılması yer almaktadır. Bu sayede NBA'nin popülerliğinin artması hedeflenmektedir. 1990'lı yıllarda Japonya'da bazı normal sezon maçlarının oynatılmasıyla başlayan bu uygulamaya daha sonra uzun bir süre ara verilmiştir. Ancak NBA yönetimi, 2010 yılında aldığı bir kararla bu uygulamayı tekrar hayata geçirmiştir. Bu sayede NBA, dünya genelindeki hayran kitlesini genişletmiştir ve uluslararası alanda daha fazla tanınmak amacıyla olan çalışmalarını sürdürmektedir (Y. Dündar, 2021: 25).

NBA'de televizyon gelirleri, eşit bir şekilde tüm takımlar arasında paylaştırılmaz. Bunun yerine, lig ve takımlar arasında bir anlaşma yapılarak belirli bir formül kullanılır. Bu formül, televizyon yayın haklarının paylaşımını belirler ve takımların başarılarına ve nüfuzlarına göre ayarlanabilir. Örneğin, daha popüler ve başarılı bir takım genellikle daha fazla televizyon geliri kazanacakken, daha az popüler veya başarılı bir takım daha az televizyon geliri kazanacaktır. Bu şekilde, her takımın işletme bütçelerini karşılayacak ve rekabeti sağlayacak bir gelir elde etmelerine olanak tanınır.

1.3. Basketbolda Oyuncu Pozisyonları

Basketbolun temel kurallarından biri, sahada aynı anda sadece 5 oyuncu oynamasıdır. Beş farklı oyuncu genellikle boylarına ve yeteneklerine göre farklı bir pozisyonlarda oynar (Xiong vd., 2017). Bu pozisyonların her birine karşılık gelen numaralarla da belirlenmiştir. Oyun kurucu 1, şutör gard 2, kısa forvet 3, güçlü forvet 4 ve merkez 5 numara olarak belirlenmiştir (Duman, 2022: 3). Bu numaraların forma numaraları ile ilişkisi bulunmamaktadır. Basketbolda forma numaraları, her takımdaki oyuncuların kimliğini belirlemek ve istatistikleri kaydetmek için kullanılır. NBA gibi profesyonel liglerde, genellikle 0'dan 99'a kadar olan sayılar kullanılabilir. Ancak, çoğu takımın belirli bir numara aralığı vardır ve bazı numaraların kullanımı takım veya lig kuralları tarafından sınırlandırılabilir.

Guard genellikle takımdaki daha kısa ve hızlı oyuncular tarafından oynanmaktadır. Genellikle en iyi dış atışları bu pozisyondakiler yapmaktadır. Sol ve sağ guard olarak içlerinde ayrılmaktadır fakat pozisyonlarında bir ayrım bulunmamaktadır. 1970’lerde belirlendiğinden

beri (Point Guard, PG) ve (Shooting Guard, SG) olarak görevleri ayrılmaktadır (Grasso, 2010: 149).

Oyun kurucu (Point Guard, PG) geleneksel olarak rolleri, topu karşı sahaya taşımak ve hücumu başlatmaktır. Genellikle takımın en kısa oyuncusu olarak yer almaktadırlar. Aynı zamanda en iyi pas verenler ve top kontrolcüleri olarak görülmektedirler (Grasso, 2010: 292).

Şutör gard (Shooting Guard, SG) oyun kurucu pozisyonuna benzese de asıl görevi hücumu organize etmek yerine bu oyunculara açılan pozisyonlar üzerinden uzak mesafeden şutları iyi değerlendirip sayı yapmalarıdır. Genellikle bir takımın en çok sayı üreten oyuncularından biri olması beklenmektedir (Grasso, 2010: 331).

Kısa forvet (Small Forward, SF) çok amaçlı oyuncu olarak hem hücumda hem de savunmada etkilidir. Genellikle orta mesafeden atıştan sorumludurlar. Bu nedenle, uzun ve iri yapılı olmaları beklenmektedir. Ancak uzun forvetten daha kısadırlar (Grasso, 2010: 332).

Uzun forvet (Power Forward, PF), olarak yer alan oyuncuların potaya yakın oynamaları ve ribaunt almaları beklendiği için genellikle uzun boylu ve güçlü olmaktadırlar. Bu oyuncuların en kötü ihtimalle orta mesafeli atışlar yapması beklenmektedir (Grasso, 2010: 300).

Pivot (Pivot Center, C) olarak adlandırılan pozisyonda yer alan oyuncular genellikle takımın en uzun oyuncularıdır. Potaya yakın oynayan bu oyunculardan hücumda pota altından basket atması, savunmada şut blok yapması beklenmektedir. Pivotların ana görevi ribaunttur (Grasso, 2010: 75).

Oyun kurucu ve şutör gard geri sahada pozisyon almaktadırlar. Forvetler ve pivot ise ön sahada oynamaktadırlar.

Tez çalışmasının dördüncü bölümünde Tablo 4.2'de veri setinde kullanılan değişkenlere ve karşılıklarına yer verilmiştir.

1.4. Win Share

NBA'de Win Share oyuncuların takımlara kazandırma paylarını gösteren resmi olmayan bir istatistiktir. Ancak bu tez çalışmasının 4.2. bölümünde yer alan Win Share değerlerinin alındığı web sitesinin kullandığı formülü olan bir istatistiktir. Birden çok NBA Win Share hesaplama formülü bulunmaktadır (Kobat, 2018: 7). Bu çalışmada kullanılan Win Share değişkeninin basketball-reference sitesinin bulduğu formüle dayanmaktadır (Basketball Reference, 2023b). Win Share değişkeni hesaplanırken kullanılan değişkenlerin de ayrıca hesaplanma formülleri bulunmaktadır. Basketball Reference sitesinin kullandığı iki Win Share değişkeni vardır. Hücum ve Savunma Win Share (Offensive Win Share, OWS ve Defensive

Win Share, DWS) değişkenlerinin formülleriyle web sitesinde yer alan en güncel hesaplama adımları şu şekildedir (Basketball Reference, 2023b). Hücum Win Share için;

1. Her oyuncu için üretilen puanları hesaplanır.
2. Her oyuncunun hücum pozisyonlarını hesaplanır.
3. Her oyuncu için marjinal hücumu hesaplanır.
4. Galibiyet başına marjinal puanları hesaplanır.
5. OWS 'i hesaplayın (Denklem 7.1).

$$OWS = \frac{Marjinal\ Hücum}{Galibiyet\ Başına\ Marjinal\ Puan} \quad (7.1)$$

Savunma Win Share için;

1. Her oyuncunun savunma derecesini hesaplanır.
2. Her oyuncu için marjinal savunmayı hesaplanır.
3. Galibiyet başına marjinal puanları hesaplanır.
4. DWS 'i hesaplayın (Denklem 7.2).

$$DWS = \frac{Marjinal\ Savunma}{Galibiyet\ Başına\ Marjinal\ Puan} \quad (7.2)$$

Win Share terimi sporcuların veya takımların bir yarışmada veya oyunun sonucunda kazandığı veya kazanmayı başardığı başarı payını ifade eder ve çalışmada çıktı değişkeni olarak kullanılmıştır (Basketball Reference, 2023b).

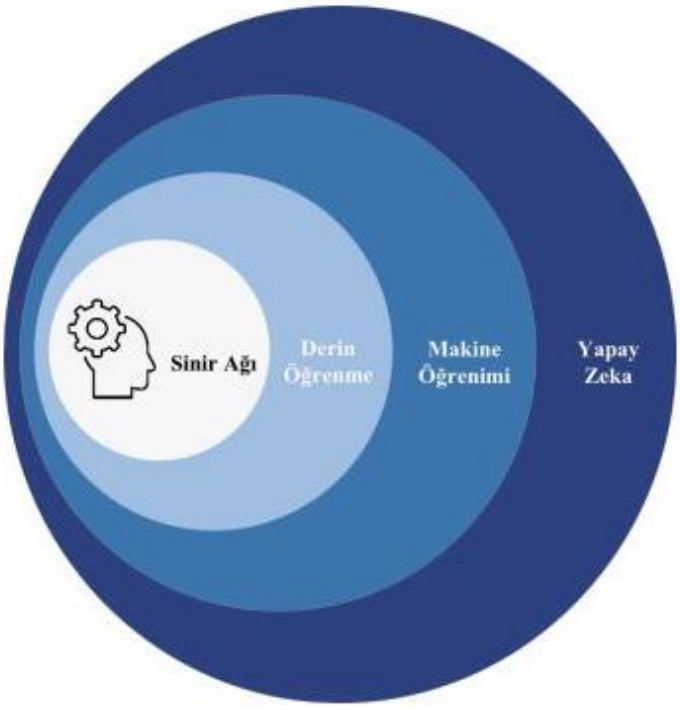
İKİNCİ BÖLÜM
MAKİNE ÖĞRENMESİ

2.1. Makine Öğrenmesi Kavramı

Makine öğrenmesi, bilgisayarların algoritmaları kullanarak öğrenmesine ve veri analizi yapmasına olanak tanıyan bir disiplindir. Makine öğrenimi, analitik model oluşturma sürecini otonom hale getirmek ve ilgili görevleri yerine getirebilmek için sistemlerin soruna özel eğitim verileri oluşturup bunları öğrenme becerisini gösterir (Janiesch vd., 2021: 685).

Makine öğrenimi, yapay zekanın alt alanı olarak kabul edilen ve belirli algoritmalar yardımıyla oluşturulan makinelerin kendi kararlarını alabilme ve sonuçlarını kullanıcılara sunabilme kabiliyetidir. Bir sistemin öğrenmesini ve gerekli kararları almasını sağlamak için algoritmaların geliştirilmesidir (Somvanshi vd., 2016: 1).

Makine öğrenmesi yapay zekânın bir alt dalı kabul edilmektedir. Derin öğrenme ise makine öğrenmesinin bir alt disiplini olarak kabul edilmektedir (Bilen, 2022: 54). Yapay zekânın makine öğrenimi ve derin öğrenme arasındaki bağlantı Şekil 2.1'de gösterilmektedir.



Şekil 2.1 Makine Öğrenmesi Yapay Zekâ Arasındaki Bağlantı

Kaynak: (Gökalp, 2022: 2)

Makine öğrenimi bilgiyi bilgisayarlara kodlamak yerine örneklerden ve gözlemlerden çıkardığı anlamlı ilişkileri ve kalıpları otomatik olarak öğrenmeye çalışır (Janiesch vd., 2021: 685).

Makine öğrenmesi disiplininin tarihçesi oldukça uzun ve köklüdür. İlk makine öğrenmesi algoritmaları, İkinci Dünya Savaşı sırasında kod çözme problemleri için kullanılan istatistiksel yöntemlerdi. İnsan zekasının makineye tanımlanabileceği ve bir makinenin insan zekasını taklit edebileceği düşünülmüştür (Cibaroğlu ve Yalçınkaya, 2019: 45).

Makine öğrenmesinin tarihçesi, 1950'lerde başlamıştır. İlk olarak, Alan Turing tarafından 1950 yılında yazılan bir makalede makine öğrenmesi kavramı ortaya atılmıştır. Alan Turing tarafından Turing testi olarak bilinen, bir makinenin insan düzeyinde zekaya sahip olup olmadığını belirlemek için kullanılan bir test önerilmiştir ve "Hesaplamalı Makineler ve Zeka" isimli makalede yayınlanmıştır (Yılmaz, 2021: 3).

1956'da Lisp dilinin geliştiricisi John McCarthy, Massachusetts Institute of Technology Yapay Zekâ Laboratuvarından aynı zamanda kurucusu da olan Marvin Minsky, Nobel ödüllü Herbert Simon, Nathaniel Rochester, ABD Yapay Zekâ Derneği'nin ilk başkanı Allen Newell ve IBM'den Claude Shannon tarafından Dartmouth Konferansı düzenlenmiştir. Yapay Zeka kavramının ilk defa kullanıldığı bu konferans, yapay zekâ alanında önemli bir kilometre taşıdır ve makine öğrenmesinin gelişiminde önemli bir rol oynamıştır (Çelebi ve İnal, 2019: 652).

1960'larda ve 1970'lerde, makine öğrenmesi alanında birçok algoritma geliştirilmiştir. Örneğin, 1960 yılında, Arthur Samuel, bir çeşit makine öğrenmesi olan "Oyunlarda Öğrenme" (Learning By Playing Games) kavramını geliştirdi. Bu kavram, programların çeşitli oyunları oynamayı öğrenmesine olanak tanıyarak, oyunlar üzerinde insanlara karşı rekabet edebilecek seviyeye ulaşmalarını sağlamaktaydı. Arthur Samuel IBM ana bilgisayar için geliştirdiği insanla karşılıklı dama oynanabilen program izin verilen hamlelerden başka bir şey bilmediği için insan karşısında yenilmekteydi ancak programı kendi kendine oynatmaya başladıktan sonra kendi hatalarından adım adım öğrenen makine insanı yenmiştir (Güney ve Yavuz, 2020: 420).

1970'lerde makine öğrenmesi araştırmaları temel olarak istatistiksel modellere dayanmaktadır ve sınırlı sayıda veri seti üzerinde çalışılmıştır. 1974 ile 1980 yılları arasında, birçok yayın yapay zeka çalışmalarını olumsuz yönde eleştirdi ve bu eleştiriler devletlerin yapay zeka çalışmalarına fon ayırmaktan vazgeçmelerine neden olduğu için bu döneme "Yapay Zeka Kışı" (Artificial Intelligence Winter) olarak adlandırılmıştır (Coşkun ve Gülleroğlu, 2021: 949).

1980'lerde yapay zekâ endüstri haline gelmiştir. 1986'da yapay sinir ağları tekrar popüler olmuştur ve yapay sinir ağlarına dayalı makine öğrenmesi yöntemleri geliştirilmiştir (Pirim, 2006: 84). Bu yöntemler, beynin çalışma prensiplerini taklit ederek, öğrenme ve veri analizi yapabilen sistemler geliştirmek için kullanılmıştır.

1990'ların başında, makine öğrenmesi teknolojileri önemli bir gelişme kaydetti ve birçok alanda kullanılmaya başlanmıştır. Özellikle, veri madenciliği ve doğal dil işleme gibi alanlarda makine öğrenmesi teknikleri başarılı bir şekilde uygulanmıştır.

1997'de önceki yıllarda IBM tarafından geliştirilen Deep Blue isimli satranç oynayabilen ilk program Kasparov ile 1-1 berabere kalmıştı ancak saniyede 200 milyon pozisyon deneyebilen ikinci versiyonu olan program dünya şampiyonu Kasparov'u yenmiştir (İpek, 2019: 206).

2000'lerin başında, büyük veri setleri ve yüksek hesaplama gücü ile makine öğrenmesi teknolojileri daha da gelişmiştir. Özellikle derin öğrenme teknikleri, yapay sinir ağlarındaki katman sayısını artırarak daha karmaşık öğrenme modelleri oluşturmayı mümkün hale getirmiştir. 2006 yılında, Geoffrey Hinton, Ruslan Salakhutdinov ve diğer araştırmacılar tarafından geliştirilen "derin öğrenme" adlı yeni bir makine öğrenmesi yaklaşımı ortaya çıktı (A. Şeker vd., 2017: 49). Bu yöntem, daha önceki yöntemlerden daha yüksek doğruluk oranları sağlayarak, görüntü, dil ve diğer alanlarda birçok başarılı uygulama yarattı.

2012 yılında, ImageNet adlı bir görüntü veri tabanı üzerinde gerçekleştirilen bir yarışmada, Alex Krizhevsky ve diğer araştırmacılar tarafından geliştirilen "Derin Sinir Ağı" adlı yeni bir makine öğrenmesi yaklaşımı, önceki yöntemlerden önemli ölçüde daha iyi sonuçlar vererek büyük bir başarı elde etti (Krizhevsky vd., 2017: 84-90). Google'ın DeepMind adlı yapay zeka şirketi 2014 yılında, atari oyunlarında insanlardan daha iyi bir performans sergiledi ve bu derin öğrenmenin uygulama alanlarını Doğal Dil İşleme (Natural Language Processing, NLP) ve konuşma sentezi gibi yeni alanlara genişletti (Ruder vd., 2019: 15). 2018 yılında, OpenAI adlı yapay zekâ şirketi, geliştirdiği bir dil modeli olan GPT-2'yi piyasaya sürdü. Aynı şirket 2020 yılında GPT-3 ü kullanıma sundu. GPT-3, önceki versiyonlarına göre daha büyük bir modeldir ve 175 milyardan fazla parametreye sahiptir (Golgiyaz, 2023: 2). Bu da onun daha geniş bir kelime dağarcığına ve daha yüksek doğruluk seviyelerine sahip olmasını sağlamaktadır. Bu model, çok sayıda NLP uygulaması için kullanılabilir ve doğal dil işleme alanındaki gelişmelere öncülük etti.

2.2. Makine Öğrenmesi Kullanılan Alanlar

Makine öğrenmesi günümüzde görüntü ve ses tanıma, doğal dil işleme, tıp ve sağlık, finans, otomotiv, pazarlama ve müşteri ilişkileri, güvenlik ve sahtekarlık tespiti, sigortacılık, enerji verimliliği, tarım, e-ticaret gibi birçok farklı alanda kullanılmaktadır.

Görüntü ve ses tanıma: Görüntü ve ses verilerini analiz etmek ve tanımak için kullanılabilmektedir. Örneğin yüz tanıma (Sharma vd., 2020: 1162-1168), nesne tanıma, el

yazısı tanıma (Preetha vd., 2020: 93-101), konuşma tanıma, görüntü sınıflandırma gibi uygulamalarda kullanılmaktadır.

Doğal dil işleme: Makine öğrenmesi, insan dilini anlamak, metinleri analiz etmek ve yorumlamak için de kullanılmaktadır. Otomatik metin çevirisi (Liang vd., 2017: 1-12), metin sınıflandırması, metin özetleme, diyalog sistemleri (Ni vd., 2022: 1-101) gibi alanlarda çalışmalar yapılmaktadır.

Tıp ve sağlık: Bu alanda tıbbi verileri analiz ederek hastalık teşhisi (Battineni vd., 2020: 1-11) ve tedavisi, ilaç keşfi (Murphy, 2011: 327-330) ve (Wale, 2011: 112-119), genetik analiz (McKinney vd., 2006: 77-88), cerrahi operasyonlar (Hashimoto vd., 2018: 70-76), radyoloji görüntüleri yorumlama (S. Wang ve Summers, 2012: 933-951) başlıklarda literatürde çalışmalar bulunmaktadır.

Finans: Finansal verileri analiz ederek hisse senedi fiyatları, piyasa trendleri, kredi riski değerlendirmesi, yatırım kararları gibi konularda tahminler yapmak için kullanılmaktadır (Çetin ve Kotan, 2023: 69). Patent analizi, online kütüphanelerde doküman eşleştirme kurumsal finans uygulamalarında kullanılmaktadır (Atalay ve Çelik, 2017: 167). Hisse senedi fiyat tahmini (Akşehir ve Kılıç, 2019: 30-39), internetten piyasa istihbaratı, acente kredi riski hesaplama (Kırca vd., 2022: 159-165), kredi riski analizi (Can, 2020: 1-57), banka verileriyle analiz (Patil ve Dharwadkar, 2017: 876-881) gibi makine öğrenmesi teknikleri kullanılarak yapılan çalışmalar literatürde yer almaktadır.

Otomotiv: Makine öğrenmesi teknikleriyle otomotiv endüstrisinde araç performansının iyileştirilmesi, otomotiv yazılımlarında (Salay vd., 2017: 1-6), otonom araç teknolojisi (Fujiyoshi vd., 2019: 244-252), sürücü yardım sistemleri, trafik kazalarını azaltmak ve şerit takibi (Chetan vd., 2020: 341-346) gibi alanlarda kullanılır.

Pazarlama ve müşteri ilişkileri: Pazarlama alanında makine öğrenmesi algoritmaları, müşteri davranışını analiz ederek kişiselleştirilmiş pazarlama stratejileri geliştirme, pazarlama çalışmalarında müşteri kaybı analizinde (Kaynar vd., 2017: 1-14), müşteri bölümleme ve davranış modellerinin belirlenmesinde (Koca, 2021: 89-111), müşteri segmentasyonu, pazarlamada fırsatlar ve pazarlama için zorlukları belirlemede (Hair Jr ve Sarstedt, 2021: 65-77), müşteri memnuniyetini artırmak ve talep tahmini gibi konularda kullanılmaktadır. Buna bir örnek olarak Yom-Tov tarafından 2018 yılında yapılan bir çalışmada, müşterilerin duygusal durumlarını otomatik olarak analiz eden bir algoritma geliştirildi. Bu sayede müşteri görüşleri ve duyguları sürekli takip edilerek, olumsuz yorumlara neden olan sorunlar belirlendi. Daha sonra ilgili firmalar tarafından yapılan düzeltmeler sonucunda, müşteri yorumları olumlu yönde

değişti. Bu çalışma sayesinde, müşteri memnuniyeti arttı ve firmalar müşteri şikayetlerine daha hızlı ve etkili bir şekilde yanıt verebileceklerini öğrenmişlerdir (Yücel, 2021: 9).

Güvenlik ve sahtekarlık tespiti: Şüpheli davranışları tespit ederek siber güvenlikte tehditleri önlemek, kredi kartı sahtekarlığı gibi konuları tespit etmek için kullanılabilmektedir. Tahminlere göre, sahtekarlık saldırıları nedeniyle tüccarların yıllık ortalama kaybı gelirlerinin %1,5'i kadar olmaktadır. Bu nedenle finansal hizmetler sektöründe risk yönetimi, yapay zekanın en önemli uygulamalarından biridir (Ahmet, 2021: 147). Finansal kuruluşlar, sahtekarlık riskini en aza indirmek ve müşterilerinin kayıplarını önlemek için çeşitli tedbirler almaktadırlar. Bu tedbirler arasında müşteri kimlik doğrulaması, sahtekarlık algılama yazılımı, şüpheli işlemlerin incelenmesi ve güvenlik protokolleri gibi bir dizi faaliyet yer almaktadır. Bu sayede finansal kuruluşlar, müşterilerinin güvenliğini sağlamak için gerekli önlemleri alarak, sahtekarlık saldırılarından kaynaklanabilecek zararları minimize edilebilir. Bu önlemler, finansal kuruluşların itibarını korurken, müşterilerin de güvenliğini sağlayarak, sektördeki rekabet avantajını da sürdürmelerine yardımcı olmaktadır.

Sigortacılık: Sigortacılık sektöründe sigorta fiyatlandırmasında (Krasheninnikova vd., 2019: 1-24), risk analizinde (Kou vd., 2019: 716-742) ,sahte hasar tahmininde (Şahin vd., 2020: 479-489) ve araç sigortası hilelerinde (Cömert ve Kaymaz, 2019: 364-390), hayat sigortası sözleşmelerinde gecikme tahmini için bir makine öğrenimi modeli (Azzone vd., 2022: 1-13) gibi çalışmalardan görüleceği gibi makine öğrenmesi kullanılmaktadır.

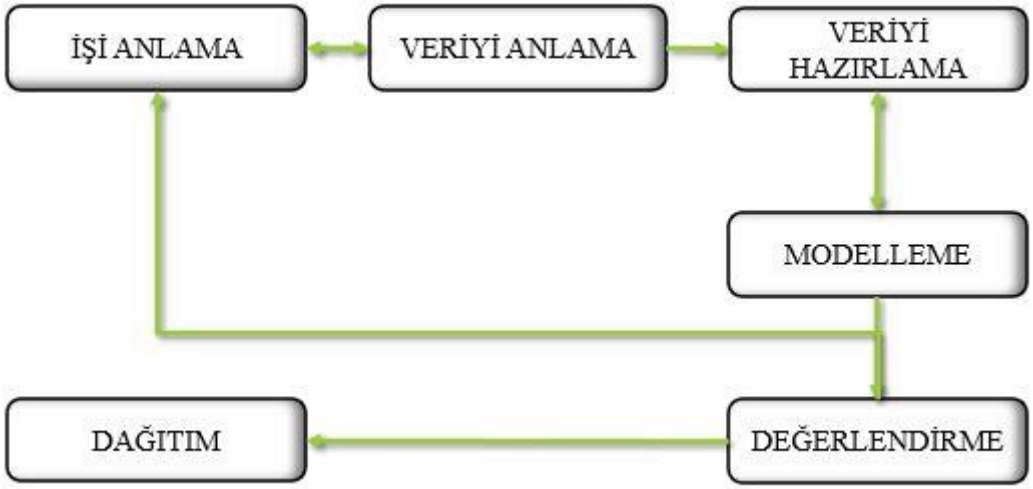
Enerji verimliliği: Enerji tüketimini analiz ederek enerji tasarrufu sağlamak (Priyadarshini vd., 2022: 1-18), elektrik talebini tahmin etmek (Bedi ve Toshniwal, 2019: 1312-1326) ve enerji sistemlerini optimize etmek için makine öğrenmesini kullanabilmek mümkündür.

Tarım: Tarım sektöründe ürün verimliliğini artırmak, ürün büyüme tahmin sistemi geliştirmek (Khan vd., 2021: 1-18), zirai hastalıkları teşhis etmek ve sulama sistemlerini optimize etmek (Abioye vd., 2022: 70-103) gibi birçok alanda kullanılabilir.

E-ticaret: Makine öğrenmesi algoritmaları, ürün önerileri sunmak, stok yönetimini optimize etmek, dijital pazarlama analitik araçlarla uygulamalarında (Miklosik vd., 2019: 85705-85718), sosyal medya pazarlamasında (Arasu vd., 2020: 1-9) ve müşteri deneyimini artırmak gibi birçok alanda kullanılabilmektedir. Örneğin Amazon gibi bir e-ticaret sitesinde, bir ürünle ilgili yüzlerce müşteri yorumu bırakılmaktadır. Bu yorumlar, metin madenciliği teknikleri kullanılarak işlenerek, ilgili ürün için özet bir tavsiye oluşturulabilmektedir (Atalay ve Çelik, 2017: 167).

2.3. CRISP-DM (Veri Madenciliği için Endüstri Standart Süreci, Cross-Industry Standard Process for Data Mining)

CRISP-DM (Cross-Industry Standard Process for Data Mining) veya Türkçe adıyla Veri Madenciliği için Endüstri Standart Süreci, veri madenciliği projelerinin yönetiminde kullanılan bir metodolojidir. CRISP-DM projesi 1997 yılının ortalarında başlamış ve kısmen Avrupa Komisyonu tarafından finanse edilmiştir (Jackson, 2002: 280). Projenin amacı, hem büyük hem de küçük veri madenciliği projelerinin gelişimini daha hızlı, daha ucuz, daha güvenilir ve daha yönetilebilir hale getirecek, endüstriden ve araçtan bağımsız bir veri madenciliği süreç modelini tanımlamak ve doğrulamaktır (Jackson, 2002: 281).



Şekil 2.2 CRISP-DM Akış Şeması

CRISP-DM, veri madenciliği projelerinin başarılı bir şekilde tamamlanması için adım adım bir rehberlik sunar. Şekil 2.2'de CRISP-DM akış şeması yer almaktadır. CRISP-DM metodolojisi, altı ana aşamadan oluşur:

İşi Anlama: Bu aşamada, veri madenciliği projesinin amacı, kapsamı ve hedefleri belirlenir. Ayrıca, projenin bütçesi, zaman çizelgesi ve kaynakları da belirlenir. Öncelikle veri madenciliği türü (örneğin sınıflandırma) ve veri madenciliği başarı kriterleri açıklanmalıdır (Schröer vd., 2021: 527).

Veriyi Anlama: Kullanılacak verilerin kaynakları belirlenir ve verilerin toplanması, incelenmesi, anlaşılması, veri kalitesini kontrol etme veriyi anlama aşamasında yapılır. Veriyi anlama işini daha somut hale getirmek için, istatistiksel analiz kullanarak ve öznitelikleri ve bunların ilişkilerini belirleyerek kullanıcı kılavuzu oluşturulur (Schröer vd., 2021: 527).

Veriyi Hazırlama: Veri hazırlamada verilerin temizlenmesi, dönüştürülmesi, birleştirilmesi ve ölçeklendirilmesi gibi işlemler yapılır. Veri madenciliği işlemlerinin daha kolay ve verimli bir şekilde yapılabilmesi için, veri tabanındaki veriler genellikle "ön işleme,

hazırlama" aşamasından geçirilir. Veri hazırlamada veriler üzerindeki gürültüleri azaltmak, veri özniteliklerini seçmek, eksik verileri tamamlamak ve gereksiz verileri atmak gibi işlemler yapılır. Bu sayede, veri madenciliği işlemleri daha az hata yapar ve daha doğru sonuçlar elde eder (Han vd., 2022: 23). Örneğin veri üzerinde eksik verilerin tespit edildiğinde, eksik verilerin tamamlanması veya eksik olan kısmın sisteme dahil edilip edilmemesi gibi kararlar verilir. Verinin tamamlanması gerektiği durumlarda, eksik kısımların tamamlanması için hangi yöntemlerin kullanılacağı belirlenir. Daha sonra alınan kararlar uygulanır ve elde edilen çıktılar değerlendirilerek daha başarılı sonuçlar elde etmek için gerekli iyileştirmeler yapılır (Ş. E. Şeker, 2018: 12). Veri hazırlama kısımlarının önceden belirlenmiş herhangi bir sırayla değil, birden çok kez gerçekleştirilme ihtimali yüksektir (Jackson, 2002: 281).

Modelleme: Modellemede, veriler üzerinde çeşitli modelleme teknikleri kullanılarak bir veya daha fazla model oluşturulur. Model oluşturulduktan sonra kullanılan tekniklerin ihtiyaçlarına uygun bir şekilde veri hazırlama aşamasına geri dönülerek gerekli değişikliklerin yapılması gerekebilir (Babaoğlu, 2015: 21). Modellemede verilerin özellikleri veya modelin performansına etki eden diğer faktörlerdeki değişiklikler, veri hazırlama aşamasında göz önünde bulundurulmalıdır. Bu sayede, modelin performansı daha iyi hale getirilir ve daha doğru sonuçlar elde edilir.

Değerlendirme: Oluşturulan modeller performanslarının ölçülmesi ve karşılaştırılması değerlendirmede yapılır. Elde edilen model (veya modeller) daha kapsamlı bir şekilde değerlendirilir ve modeli oluşturmak için yürütülen adımlar, iş hedeflerine uygun şekilde ulaştığından emin olmak için gözden geçirilir (Azevedo ve Santos, 2008: 184). Değerlendirme aşamasında önemli bir diğer amaç ise, yeterince dikkate alınmamış bazı önemli iş konularının olup olmadığını belirlemektir (Wirth ve Hipp, 2000: 35).

Dağıtım: Model oluşturmak genellikle bir projenin son adımı değildir. Modelden elde edilen bilgilerin, müşterinin kullanabileceği şekilde düzenlenip sunulması gerekebilmektedir. Bu durum dağıtım aşaması olarak adlandırılır ve bir rapor oluşturmak kadar basit olabilmektedir. Aynı zamanda sürekli tekrarlanabilen bir veri madenciliği süreci uygulamak kadar karmaşık da olabilmektedir. Dağıtımda, genellikle kullanıcılar, veri analisti yerine, oluşturulan modellerden faydalanacak olan kişilerdir. Bu nedenle, hangi eylemlerin gerçekleştirilmesi gerektiğini önceden anlamak, elde edilen sonuçların gerçekten faydalı olmasını sağlar (Wirth ve Hipp, 2000: 36). Bu son aşama, modelin kullanıma hazır hale getirildiği ve organizasyonda uygulandığı kısımdır. Modelin etkin bir şekilde kullanılması ve sonuçların yorumlanması için gerekli önlemler alınır. Modelin gerçek zamanlı uygulanması ve sonuçlarının izlenmesi sağlanır.

CRISP-DM, veri madenciliği projelerinin her adımını yönetmek için detaylı bir planlama ve yönetim süreci sunar. Bu sayede, veri madenciliği projeleri başarıyla tamamlanabilir ve işletmelere değerli bilgiler sunulabilir.

"CRISP-DM Twenty Years Later: From Data Mining Processes to Data Science Trajectories" isimli çalışmalarında Martinez-Plumed ve diğerlerine göre CRISP-DM, veri madenciliği projelerini kurmak ve yönetmek için ortak bir çerçeve olarak hala önemli bir rol oynamaktadır (Martínez-Plumed vd., 2019: 3058). Bu tez çalışmasında CRISP-DM metodu kullanılmıştır.

Makine öğrenmesi veri madenciliği alanında kullanılan bir dizi algoritma ve teknikleri içeren bir disiplindir. Bu nedenle CRISP-DM makine öğrenmesi projelerini yönetmek için kullanılabilecek bir çerçeve sunmaktadır. CRISP-DM bir makine öğrenmesi projesini adımlara bölerek yönetilebilir hale getirir ve projenin başarılı bir şekilde tamamlanmasına rehberlik etmektedir.

Veri madenciliği sürecinin makine öğrenmesine evrilmesi veri madenciliği alanının gelişimiyle birlikte gerçekleşmiştir. İlk başlarda veri madenciliği, verilerin keşfedilmesi ve analiz edilmesiyle ilgili bir disiplin olarak gösterilmiştir (Ersöz ve Çınar, 2021: 402). İstatistiksel yöntemler ve veri görselleştirme araçları kullanılarak verilerin içindeki desenler ve ilişkiler ortaya çıkarılmaya çalışılmıştır (Erden, 2021: 6).

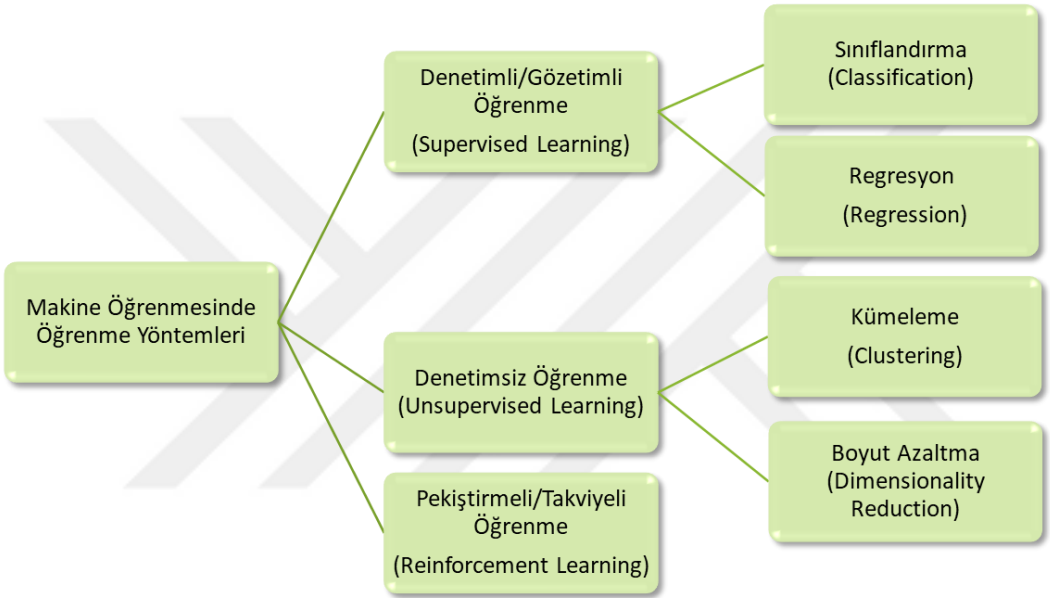
Ancak teknoloji ilerledikçe ve bilgisayarların gücü arttıkça, veri madenciliği süreci daha karmaşık hale gelmiştir. Büyük veri setleriyle çalışmak ve bu verilerden anlamlı sonuçlar çıkarmak için daha gelişmiş analitik yöntemlere ihtiyaç duyulmuştur. Makine öğrenmesi algoritmaların ve istatistiksel modellerin kullanıldığı bir disiplindir. Bu alanda kullanılan algoritmalar, bilgisayarların verilerden öğrenme yeteneğine dayanmaktadır. Makine öğrenmesi modelleri büyük veri setlerinden desenleri tespit etmeyi, tahminlerde bulunmayı ve kararlar almaya yardımcı olmayı amaçlamaktadır (Carbonell vd., 1983: 8).

Veri madenciliği süreci içinde makine öğrenmesi, veri hazırlama ve modelleme aşamalarında önemli bir rol oynar. Veri hazırlama aşamasında, makine öğrenmesi için uygun bir formatta veriler hazırlanır ve özellik seçimi veya öznitelik mühendisliği gibi işlemler gerçekleştirilir. Modelleme aşamasında ise, makine öğrenmesi algoritmaları kullanılarak veriler üzerinde eğitim yapılır ve bir model oluşturulur. Makine öğrenmesi, veri madenciliği sürecini daha otomatik ve veri odaklı hale getirirken, daha karmaşık analizlerin yapılabilmesini sağlar. Büyük veri setleri üzerindeki desenleri keşfetmek, tahminlerde bulunmak veya sınıflandırma yapmak gibi görevleri gerçekleştirebilir. Makine öğrenmesi otomasyon, yapay zeka ve veri bilimi alanlarının da temelini oluşturur (Erden, 2021: 2).

Makine öğrenmesi veri madenciliği tekniklerinin yanında grafikleri, doğal dil işleme tekniklerini, yapay sinir ağları gibi farklı öğrenme yöntemlerini kullanır (Ersöz ve Çınar, 2021: 402). Bir sonraki bölümde makine öğrenmesinde kullanılan öğrenme yöntemlerinden bahsedilecektir.

2.4. Makine Öğrenmesinde Öğrenme Yöntemleri

Makine öğrenmesinde öğrenme; denetimli, denetimsiz ve pekiştirmeli (takviyeli) olmak üzere üç grupta incelenmektedir. Makine öğrenmesinde öğrenme grupları ve modellerinin şemasına Şekil 2.3'te yer verilmiştir.



Şekil 2.3 Makine Öğrenmesi Yöntemleri

2.4.1. Denetimli/Gözetimli Öğrenme (Supervised Learning)

Denetimli öğrenme, bir modelin belirli bir etiketle eşleştirilmiş verilerle eğitildiği bir öğrenme yöntemidir. Bahsedilen etiket eğitim veri kümesindeki doğru cevap olarak bilinmektedir. Modelin girdi değerleriyle etiketlenmiş değerler arasındaki ilişkiyi öğrenmesi ve hedef değerlere en yakın çıktıların bulunması amaçlanmaktadır (Atalay ve Çelik, 2017: 161). Denetimli öğrenme kavramı bazı çalışmalarda, "gözetimli öğrenme" olarak da yer almaktadır. Girdi değişkenlerine bağımsız değişken, çıktı değişkenlerine hedef değişken de denmektedir (Kırca vd., 2022: 161).

Makine öğrenmesi için kullanılacak veriler eğitim verisi (training data) ve test verisi (test data) olarak ikiye ayrılır. Kullanılacak algoritma için gözlemlerden oluşan veriler eğitim verisi olarak kullanılmaktadır. Eğitim aşamasında eğitim için ayrılan verilerle ele alınan algoritma ile çıkarımlarda bulunulur ve bir model oluşturulup, modelin parametreleri tahmin edilir (Akay, 2018: 46). Test aşamasında ise test verileriyle yeni tahminlenerek üretilecek

verilerin tahmininin doğru olup olmadığına dair performans kriterlerine göre hesaplamalar yapılmaktadır (Altın Yavuz ve Yavuz, 2021: 143). Tahmin edilen çıktıları gerçek çıktılarla yani etiketlerle karşılaştırarak bir hata hesaplanır ve bu hata, modelin parametrelerinin güncellenmesi için kullanılmaktadır. Bu işlem, modelin hedef çıktılara en yakın sonuçları üretmesi için parametrelerini ayarlamasına yardımcı olmaktadır. Yeni bir girdi olduğunda en son oluşan model en yakın çıktıyı verecektir (Atalay ve Çelik, 2017: 161).

Bu tür öğrenme algoritmaları, birçok gerçek dünya uygulamasında sıklıkla kullanılır. Örneğin, bir e-postanın spam mı yoksa spam değil mi olduğunu belirlemek, bir resimdeki nesneleri tanımak veya bir müşterinin bir ürünü satın alma olasılığını tahmin etmek, bir hastanın hastalığını tahmin etmek veya bir tedavi yönteminin sonucunu tahmin etmek denetimsiz makine öğrenmesidir.

Denetimli öğrenme modeller geliştirmek için sınıflandırma ve regresyon yöntemi olarak ikiye ayrılır (Kırca vd., 2022: 160).

2.4.1.1. Sınıflandırma (Classification)

Sınıflandırma, makine öğrenmesinde yaygın olarak kullanılan bir yöntemdir. Sınıflandırma yönteminin amacı verilen veri özelliklerine dayanarak belirli bir sınıfa ait olduğu düşünülen bir nesnenin sınıfını belirlemektir ve var olan verilerden örüntü keşfetmektir (Atalay ve Çelik, 2017: 165). Denetimli öğrenme problemlerinin bir alt kümesidir ve etiketli veri setleri kullanır. Sınıflandırma, özellikleri verilen bir nesneyi tanımlamak ve bu nesneyi belirli bir sınıfa atamak için kullanılmaktadır. Her sınıfın, sınıf içinde yer alan tüm nesnelerin paylaştığı belirli bir ortak özelliği veya özellikleri vardır (Babaoğlu, 2015: 25). Veri seti etiketlenebiliyor, kategorileşebiliyor ya da belirli gruplara ayrılabiliyorsa, sınıflandırma teknikleri kullanılabilir (Can, 2020: 5).

Sınıflandırma iki ana sınıf üzerinde yapılabilir: İki sınıflı sınıflandırma ve çok sınıflı sınıflandırma. İki sınıflı sınıflandırma, bir nesnenin iki farklı sınıftan birine ait olduğu durumlarda kullanılır. Örneğin, bir e-postanın spam veya spam olmayan klasörüne yerleştirilmesi gibi. Çok sınıflı sınıflandırma ise, nesnelerin birden fazla sınıf altında sınıflandırılabileceği durumlarda kullanılır. Fasulye tohumlarının sınıflandırılması buna örnek olarak verilebilir (Bülbül, 2023: 33-43).

Sınıflandırma yöntemleri arasında, Karar Ağaçları, K-NN (K-En Yakın Komşu Algoritması) Sınıflandırıcısı, Destek Vektör Makineleri (Support Vector Machines, SVM), Yapay Sinir Ağları (YSA), Naive Bayes Sınıflandırıcısı, Lineer Diskriminant Analizi, Lojistik Regresyon, Genetik Algoritmalar gibi birçok farklı yöntem bulunur (Güvenç vd., 2022: 1360).

Bu yöntemler, veri özelliklerine, veri boyutuna, etiket sayısına ve veri özelliklerinin dağılımına bağlı olarak farklı seviyelerde etkilidir. Veriye uygun olarak algoritmalar seçilir.

2.4.1.2. Regresyon (Regression)

Makine öğrenmesinde regresyon, bir ya da birden fazla bağımsız değişkenle bağımlı değişken arasındaki ilişkiyi modeller. Bu ilişki, bir bağımlı değişkenin, bir veya daha fazla bağımsız değişkene etkisini tahmin etmek için kullanılır.

Bir bağımlı değişken ve ikiden fazla bağımsız değişkenin yer aldığı regresyon yöntemleri çoklu regresyon olarak adlandırılmaktadır (Akşehir ve Kılıç, 2019: 33). Regresyon, sürekli bir çıktı değişkeni tahmin etmek için kullanılır. Örneğin, evlerin satış fiyatını, hava durumu verilerini veya hisse senedi fiyatlarını tahmin etmek için regresyon kullanılabilir.

Sınıflama, verilerin belirli kategorik sınıflara ayrıldığı bir tahminleme yöntemidir ve kategorileri tahmin etmektedir. Regresyon ise, veriler arasındaki sürekli bir ilişkiyi modeller ve süreklilik gösteren değerleri tahmin etmektedir (Babaoğlu, 2015: 25). Regresyon yöntemleri sürekli değişim gösteren durumlardaki verilerin çıktı değişkenlerini öngörmektedir (Can, 2020: 5). Regresyon yöntemleri arasında Karar Ağaçları, Destek Vektör Makineleri (Support Vector Machines, SVM), Yapay Sinir Ağları (YSA), Lineer Regresyon, Rastgele Orman Regresyon, Gradient Boosting Regresyon ve XGBoost Regresyon gibi algoritmalar vardır (Kırca vd., 2022: 162).

2.4.2. Denetimsiz Öğrenme (Unsupervised Learning)

Denetimsiz öğrenme, öğrenme veri kümesinde sınıflandırılmamış veya kategorize edilmemiş veri kümelerine herhangi bir etiketleme olmadan verilerin işlenmesi ve yapılandırılmasıdır (Ahmet, 2021: 152). Daha çok kümeleme, olasılık yoğunluk tahmini, öznitelikler arasındaki ilişkilerin bulunması ve boyut indirgeme gibi sebeplerle denetimsiz öğrenme uygulanmaktadır (Bilgin, 2017: 2). Eğitim sırasında etiketlenmemiş veriler kullanılarak bu içgörüleri, ilişkileri anlamak, tanımak ve keşfetmek için kullanılır. Bu nedenle, modelden belirli bir sonuç beklemek mümkün değildir. Bu model doğru cevaplarla eğitilmemiştir, ancak gözetimsiz, kendi başına kalıpları ve desenleri bulması beklenir (Gökalp, 2022: 6). İnsanlar verileri sürekli sınıflandırdığı için sınıflandırma, hem makine öğrenmesinin temeli hem de veri hazırlama aracı olarak da kullanabilmektedir (Akın, 2008: 56).

Bu tür öğrenme algoritmaları, birçok farklı alan için kullanılabilir. Bir müşteri veri kümesindeki benzer müşteri gruplarını tanımlamak, bir web sayfasının farklı bölümlerini belirlemek veya bir resimdeki benzer pikselleri kümelemek denetimsiz öğrenmeye örnek

verilebilir. Denetimsiz öğrenme kümeleme ve boyut azaltma yöntemi olarak ikiye ayrılmaktadır (Kırca vd., 2022: 160).

2.4.2.1. Kümeleme (Clustering)

Kümeleme, heterojen olan ana kütleleri homojen gruplara veya kümeler halinde ayırmayı amaçlar. Kümeleme önceden tanımlanmış sınıfları baz alarak gruplama yapmaz; bunun yerine benzerlikleri kullanarak gruplandırma işlemi gerçekleştirir. Kümeleme yönteminin ve sınıflandırmanın farkı ise sınıflandırma yönteminde önceden tanımlanmış sınıf ya da kategori modeline dayanarak verileri sınıflandırır (Akın, 2008: 56).

Kümeleme yöntemi farklı alanlarda birçok uygulamada kullanılabilir. Örneğin, müşteri segmentasyonu, genetik veri analizi, pazarlama stratejileri, sosyal medya analizi, tıbbi teşhisler ve daha birçok alanda bu yöntemden yararlanılabilir.

2.4.2.2. Boyut Azaltma (Dimensionality Reduction)

Denetimsiz öğrenmede boyut azaltma, yüksek boyutlu veri setinin orijinal verilerdeki bilgiler korunurken boyutunu azaltarak, verilerin daha anlamlı hale getirilmesini amaçlayan bir yöntemdir (Demirci Gülpınar, 2022: 1458). Bu yöntem, verilerin içindeki gürültüleri ve gereksiz özellikleri ayıklar ve geriye kalan özelliklerle daha az boyutlu bir veri seti oluşturur.

Boyut azaltma yöntemleri, verilerin yapısal özelliklerini korurken, boyutları azaltarak veri setinin işlenmesini ve analiz edilmesini kolaylaştırır. Bu yöntemler sayesinde, veri setleri daha basit ve anlaşılır hale getirilir ve daha az bellek kullanımı sağlanır. Boyut azaltma, bilgi kaybını en aza indirir ve verilerde bulunan girdi değişkenlerin sayısını azaltmaktadır (D. R. Dündar vd., 2021: 264).

Boyut azaltma yöntemleri, veri setleri içindeki özelliklerin daha iyi anlaşılmasına yardımcı olur ve veri setleri üzerinde daha verimli işlemler yapılmasını sağlar.

2.4.3. Pekiştirmeli/Takviyeli Öğrenme (Reinforcement Learning)

Pekiştirmeli (takviyeli) öğrenme, bir modelin bir çevreyle etkileşime girdiği ve bir ödül veya ceza aldığı bir öğrenme yöntemidir (Mousavi vd., 2018: 2). Pekiştirmeli öğrenme, genellikle araç, ortam, ödül ve eylem olmak üzere dört temel bileşenden oluşmaktadır (Sağ ve Hasırcı Tuğcu, 2022: 42). Bir ajanın bir ortamla etkileşim halinde olduğu bu yaklaşımda, ajan belirli bir durumda çevreden bir gözlem yapar ve bu gözlemi işleyerek bir eylem gerçekleştirir. Bu eylem, çevreyi etkiler ve ajan yeni bir duruma geçer. Ortam, ajanın yaptığı eyleme bağlı olarak bir ödül verir veya ceza uygular. Ajanın amacı, ortamdan aldığı geri bildirimleri

kullanarak öğrenmek ve ödülü en üst düzeye çıkarmak için en iyi eylemleri seçmektir (Sağ ve Hasırcı Tuğcu, 2022: 42).

Bu tür öğrenme algoritmaları özellikle robotik, oyunlar ve diğer benzer uygulamalar için kullanılır. Örnek olarak bir robotun bir nesneyi almasını öğrenmek, bir bilgisayar oyunundaki karakterin doğru kararları almasını sağlamak veya bir araç sürerken doğru kararları almak verilebilir. Son on yılda pekiştirmeli öğrenme yöntemiyle, kişiselleştirilmiş tıpta etkili uygulamalar bulunmaktadır (Coronato vd., 2020: 1).

2.5. Makine Öğrenmesi Algoritmaları

Bu tez çalışmasında veri seti ön işleme adımlarından sonra veri seti öğrenme ve test olarak ikiye ayrılmıştır. Modelleme aşamasında veri seti üzerinde yapılan düzenlemelerin ardından hangi makine öğrenimi algoritmalarının eğitim için kullanılacağı belirlenmiş ve eğitim modeli oluşturulmuştur. Gerçekleştirilen tez çalışmasının modelleme aşamasında verisetine uygun olarak Denetimli/Gözetimli Öğrenme (Supervised Learning) yöntemlerinden Regresyon (Regression) modelleri kullanılmıştır. Literatürde bahsedilen makine öğrenmesi algoritmalarından Yapay Sinir Ağları (Neural Network), Karar Ağaçları (Decision Tree), Karar Ağaçları Topluluk Öğrenmesi (Decision Tree Ensemble), Rastgele Orman (Random Forest), Lineer Regresyon (Linear Regression) ve XGBoost Learning algoritmaları seçilmiştir ve bu modeller eğitimde kullanılmıştır. Devam eden bölümde kullanılan algoritmaların detaylı açıklamalarına yer verilmiştir.

2.5.1. Yapay Sinir Ağları (Neural Network)

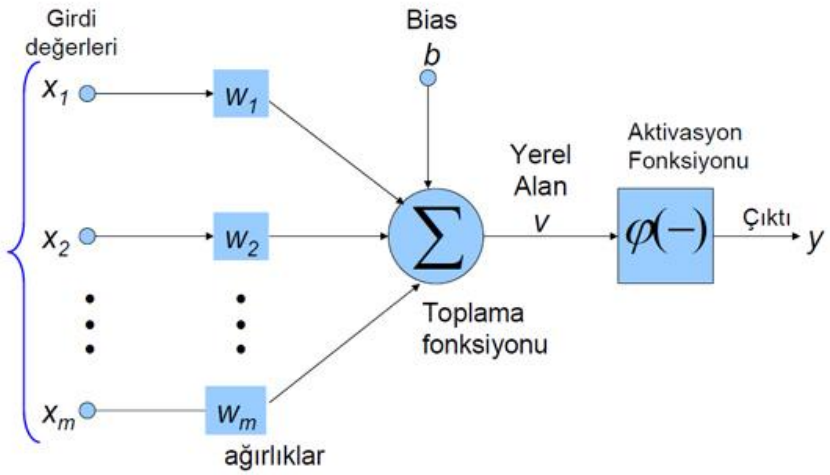
Yapay sinir ağları biyolojik sinir sistemlerini taklit ederek, veriler arasındaki karmaşık ilişkileri öğrenmek ve tahmin etmek için kullanılan bir makine öğrenmesi yöntemidir (Sargent, 2001: 1640).

Tablo 2.1 Biyolojik Sinir Sistemi Elemanları ve Yapay Sinir Sisteminde Karşılıkları

Biyolojik Sinir Sistemi	Yapay Sinir Ağları
Nöron	İşlemci Elemanı
Dentrit	Toplama Fonksiyonu
Hücre Gövdesi	Aktivasyon Fonksiyonu
Akson	Çıkış
Sinaps	Ağırlık

Yapay sinir ağları insandaki biyolojik sinir sistemlerinden esinlenerek tasarlanmıştır. Tablo 2.1'de biyolojik sinir sisteminin elemanları ve yapay sinir sisteminde karşılıkları gösterilmiştir. Biyolojik sinir hücreleri, sinaps adı verilen bağlantılar vasıtasıyla birbirleriyle iletişime geçerler. Bir sinir hücresi, işlediği bilgileri aksonları yoluyla diğer hücrelere gönderir.

Yapay sinir ağları da benzer şekilde çalışır. Ağın her bir hücresi, diğer hücrelerle bağlantılar kurarak bilgiyi işler. Bu bağlantılar, ağırlık ve bias değerleri gibi parametrelerle belirlenir ve ağın öğrenmesinde büyük bir rol oynar. Her hücre, girdileri toplama ve aktivasyon fonksiyonları kullanarak işler ve çıktıyı diğer hücrelere aktarır. Yapay sinir ağlarında kullanılan toplama ve aktivasyon fonksiyonları çeşitli matematiksel işlemlerden oluşur. Toplama fonksiyonu, girdileri toplar ve ağırlıklarla çarparak ağırlıklı toplamlarını hesaplar. Aktivasyon fonksiyonu ise toplamın bir eşik değerden büyük veya küçük olup olmadığını kontrol eder ve buna göre çıktıyı belirler. Farklı toplama ve aktivasyon fonksiyonları bulunmaktadır (Öztemel, 2003: 45). Şekil 2.4'de yapay sinir ağları yapısı gösterilmiştir.



Şekil 2.4 Yapay Sinir Hücresi Yapısı

Kaynak: (Çınar, 2018)

Yapay sinir ağları; girdi katmanı, gizli katmanlar ve çıktı katmanı olmak üzere üç temel katmandan oluşur. Girdi katmanı, verilerin alındığı katmandır. Bu katmanda herhangi bir bilgi işleme yapılmaz. Gizli katman(lar) ise, girdileri işlemek, öğrenmek ve bilgiyi özetlemek için kullanılan katmanlardır. Genellikle tek ara katmanla çoğu problemi çözebilmek mümkündür. Ancak ağın öğrenmesi istenen sorunun girdi ve çıktı arasındaki ilişki doğrusal değilse ve karmaşıklığın arttığı durumda çok sayıda gizli katman kullanılabilmektedir. Son olarak çıktı katmanı nihai sonuçların üretildiği katmandır (Öztemel, 2003: 75).

Her bir katman bir dizi yapay sinir hücresinden oluşur. Yapay sinir hücreleri ağırlık değerleri ve aktivasyon fonksiyonları içeren matematiksel işlemleri gerçekleştirir. Ağırlıklar sinir ağının öğrenme süreci sırasında güncellenir. Bu işlemler sinir ağı tarafından bir modele dönüştürülür ve çıktı katmanından sonuçlar elde edilir (Şen, 2004). Yapay sinir ağlarının öğrenme süreci, genellikle backpropagation (geriye yayılım) adı verilen bir yöntemle gerçekleştirilir. Geriye Yayılım yöntemi sinir ağının ürettiği çıktılarla hedef çıktılar arasındaki farkı hesaplar ve bu farkı kullanarak ağdaki ağırlık değerlerini günceller. Bu işlem birkaç kez

tekrarlanarak sinir ağının doğruluğu artırılır. Yapay sinir ağları, özellikle karmaşık veri yapıları üzerinde işlem yapmak için kullanılır. Örneğin, görüntü ve ses tanıma, doğal dil işleme, finansal tahminler ve tıbbi teşhis gibi alanlarda başarıyla kullanılmaktadır. Ancak yapay sinir ağları modellerinin karmaşıklığı nedeniyle, eğitim ve hiperparametre ayarlaması için yeterli veri, hesaplama gücü ve zaman gerektirirler (Mehonic vd., 2020: 7).

Yapay Sinir Ağları, birçok hiperparametreye sahip olan karmaşık bir modeldir. Hiperparametreler, ağın yapısını ve eğitim sürecini kontrol etmek için kullanılır (Diaz vd., 2017: 1). Bu hiperparametrelerin doğru ayarlanması, ağın performansını etkileyebilir. Yapay sinir ağlarında Öğrenme Oranı (Learning Rate), Epok Sayısı (Number of Epochs), Gizli Katman Sayısı ve Boyutu (Number and Size of Hidden Layers), Aktivasyon Fonksiyonu (Activation Function) ve Batch Boyutu (Batch Size) hiperparametreleri yaygın olarak kullanılmaktadır (Aszemi ve Dominic, 2019: 271).

Öğrenme Oranı (Learning Rate) ağı güncellemek için kullanılan adım boyutunu belirler. Bu değer her ağırlık güncellemesinde ne kadar ilerleme kaydedileceğini belirlemektedir. Öğrenme oranı çok küçükse ağ yavaş bir şekilde öğrenirken, çok büyükse ağın yakınsama sorunları yaşamasına neden olabilir (Zeiler, 2012: 1).

Epok Sayısı (Number of Epochs) ağın eğitim verilerini kaç kez göreceğini belirler. Ağın daha fazla epokla eğitilmesi, daha fazla öğrenme sağlayabilir ancak aşırı öğrenmeye de yol açabilir (Brownlee, 2018: 4).

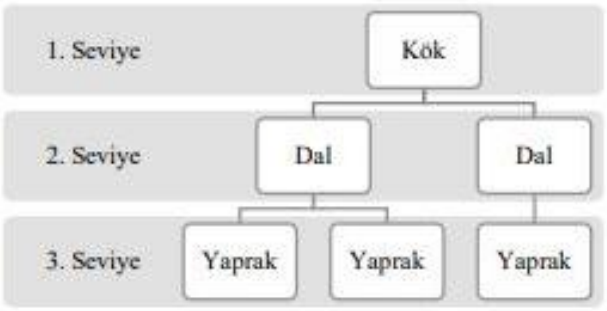
Gizli Katman Sayısı ve Boyutu (Number and Size of Hidden Layers): Yapay sinir ağlarında verilerin işlenmesi için bir veya daha fazla gizli katman kullanılabilir. Gizli katmanlar ağın karmaşıklığını ve temsil yeteneğini artırır (Smithson vd., 2016: 2).

Aktivasyon Fonksiyonu (Activation Function) her sinir hücresinin çıkışını belirler. Bu fonksiyon ağın Lineer olmayan (non-lineer) ilişkileri öğrenebilmesini sağlar. Sigmoid, ReLU, tanh gibi farklı aktivasyon fonksiyonları kullanılabilir (Dubey vd., 2022: 3). Hangi aktivasyon fonksiyonunun kullanılacağı da bir hiperparametredir.

Batch Boyutu (Batch Size) eğitim verilerinin ağa hangi gruplar halinde verileceğini belirler. Büyük bir batch boyutu daha hızlı eğitim sağlayabilirken, bellek gereksinimlerini artırır. Küçük bir batch boyutu daha yavaş eğitim süresine neden olabilir ancak daha iyi genelleştirme yapabilir (Brownlee, 2018: 3). Ağ performansında etkileyebilen hiperparametreler genellikle deneme yanılma yoluyla ayarlanır. Deneyerek ve sonuçları değerlendirerek, en uygun hiperparametre değerlerini bulmak önemlidir.

2.5.2. Karar Ağacı (Decision Tree)

Karar ağaçları, ağaç yapısı şeklinde bir model oluşturarak verileri sınıflandırır veya regresyon analizi yapar. Karar ağacı en üstteki eleman olan kök düğümle başlar ve her düğüm belirli bir özelliğin bir değerine göre alt düğümlere ayrılır. En alttaki elemana yaprak ve arasındaki kalan elemanlara da dal denmektedir. Bu işlem, tüm veri seti için tekrarlanarak ağacın yaprak düğümlerine kadar ilerlenir (Daş ve Türkoğlu, 2014: 382). Yaprak düğümler, sonuçların verildiği ve sınıflandırma veya regresyon sonuçlarını veren düğümlerdir. Şekil 2.5'de bir karar ağacı yapısı gösterilmektedir.



Şekil 2.5 Karar Ağacı Yapısı

Kaynak: (Daş ve Türkoğlu, 2014: 382)

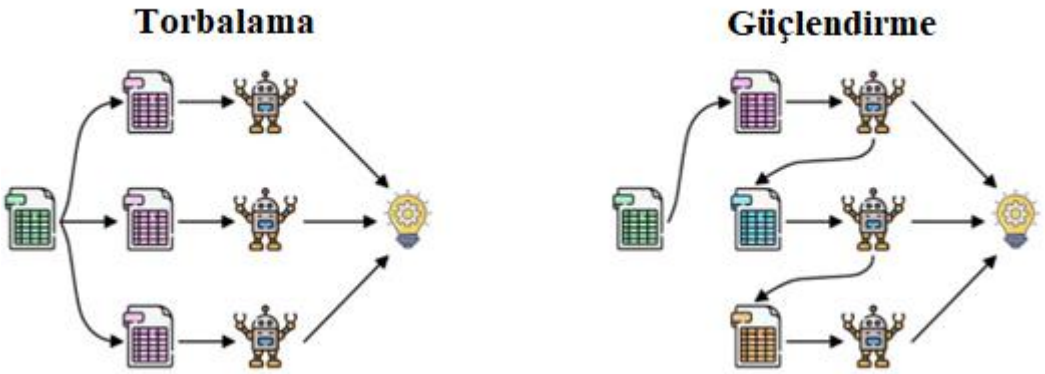
Veriler, basit karar verme adımları uygulanarak, belirli özelliklere göre alt gruplara bölünür. Her başarılı bölme işlemi sonucunda, oluşan gruplar birbirine daha benzer hale gelir. Karar ağaçları, sınıflandırma ve regresyon problemlerinde yüksek doğruluk oranları sağlar ve ayrıştırıcı (discriminatory) özellikleri tespit etmede başarılıdır. Diğer sınıflandırma ve regresyon yöntemlerinden daha kolay anlaşılır ve yorumlanabilir sonuçlar vermesi nedeniyle yaygın olarak kullanılmaktadır (Albayrak ve Yilmaz, 2009: 39). Karar ağaçları tümevarım yöntemiyle hareket eder (Taşkın ve Gökay, 2005: 225).

Karar ağaçlarında en çok karşılaşılan sorun overfitting (aşırı öğrenme)'dir. Bu durum karar ağacının eğitim veri setiyle öğrenirken aşırı genelleme yapmasıdır. Veri setini ezberleyen karar ağaca prunning (budama) işlemi uygulanabilmektedir (Koçak, 2020: 75).

2.5.3. Karar Ağaçları Topluluk Öğrenmesi (Decision Tree Ensemble)

Karar Ağaçları Topluluk Öğrenmesi (Decision Tree Ensemble), bir araya getirme, Topluluk Öğrenme (Ensemble) yöntemidir. Bu yöntemde birçok karar ağacı oluşturulur ve sonuçlar bir araya getirilerek bir ortak tahmin elde edilir. Birçok karar ağacının bir araya gelmesi, tek bir karar ağacından daha yüksek tahmin performansı elde etmek için kullanılır. Topluluk Öğrenmesi modelinin temel prensibi, zayıf öğrenicilerin birleşerek güçlü bir öğrenici

oluşturmasıdır (Nagpal, 2017). Topluluk Öğrenmesi, karar ağaçlarını gerçekleştirmek için Torbalama (Bagging) ve Güçlendirme (Boosting) gibi iki yöntem kullanılmaktadır.



Şekil 2.6 Torbalama ve Güçlendirme Yöntemleri

Kaynak: (López, 2021)

Torbalama (Bagging), "Bootstrap AGGregatING"'in kısaltılmış halidir (Breiman, 1996: 123). Torbalama birçok karar ağacını bağımsız olarak oluşturur ve bu ağaçların sonuçlarını birleştirerek bir ortak tahmin elde eder. Her ağaç rastgele bir alt küme seçilen örnekler kullanılarak bağımsız olarak büyütülür. Bu nedenle Torbalama yöntemi ağaçların varyansını azaltmak ve daha istikrarlı bir tahminleme performansı elde etmek için kullanılır (Lemmens ve Croux, 2006: 278).

Güçlendirme (Boosting) yöntemi birçok zayıf öğreniciden güçlü bir öğrenici oluşturmak için kullanılır. Bu yöntemde her ağaç önceki ağaçların hatalarına göre ağırlıklı bir şekilde oluşturulur. Bir ağaç diğer ağaçların yanlış sınıflandırılan örneklerine odaklanır ve bu örneklerin sınıflandırmasını doğru bir şekilde yapmaya çalışır. Güçlendirme yöntemi, Torbalama yönteminden daha yüksek bir doğruluk elde etmek için kullanılır (Nagpal, 2017). Şekil 2.6'da Torbalama ve Güçlendirme yöntemlerinin gösterimi yer almaktadır.

Karar Ağaçlarına dayalı en popüler topluluk öğrenmesi yöntemlerinden bazıları şunlardır:

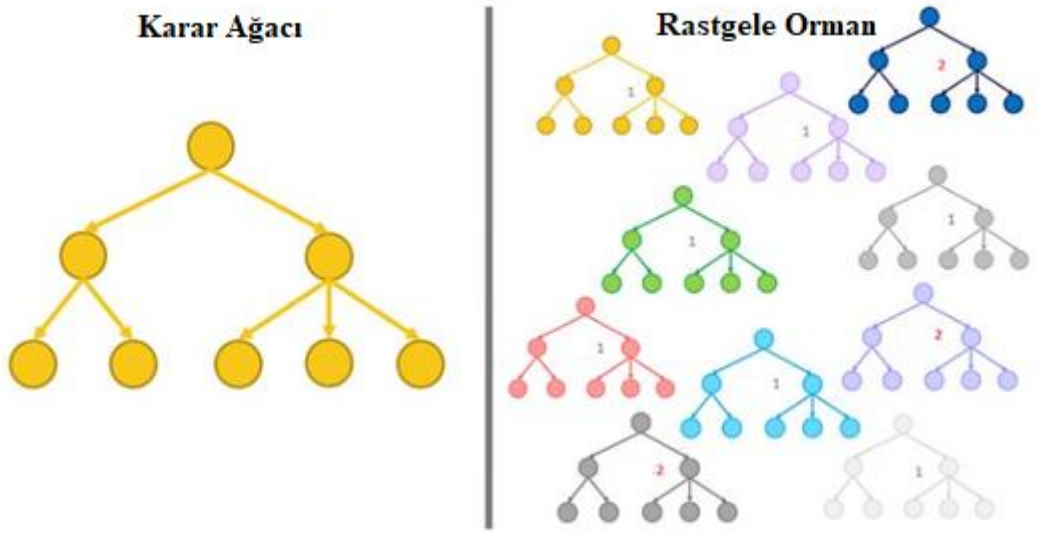
- Rastgele Orman
- Extremely Randomized Trees
- Torbalama
- Adaptive Booster
- Gradient Boost
- XGBoost

Bu topluluk yöntemlerinin tümü bir karar ağacı alır ve ardından varyansı ve yanlılığı azaltmanın bir yolu olarak ya torbalama ya da güçlendirme uygular (Slatery, 2020). Bu yöntemlerin hepsi hem regresyon hem de sınıflama için kullanılmaktadır.

2.5.4. Rastgele Orman (Random Forest)

Rastgele Orman Torbalama (Bagging)'in bir uzantısıdır. Rastgele veri alt kümesi alır ve ağaçları büyütmek için tüm özellikleri kullanmak yerine özelliklerin rastgele seçimini de yapar. Birçok rastgele ağaca sahip olduğunda algoritma Rastgele Orman olarak adlandırılır (Nagpal, 2017). Karar ağacı oluştururken, her seviyedeki öznitelik belirleme işlemi, belirli bir kural veya kriter doğrultusunda gerçekleştirilir. Bu kural veya kriter, veri setindeki değişkenlerin belirleyici önemlerine göre belirlenir. Belirli bir öznitelik, veri setindeki en iyi ayırıcı özelliklerden biri olarak kabul edildiğinde, o öznitelik seçilir ve ağaca dahil edilir. Diğer seviyelerde de aynı işlem tekrarlanır ve sonunda bir sonuç elde edilir (Daş ve Türkoğlu, 2014: 382). Birkaç Rastgele Orman ağacını birleştiren ve tahminlerini ortalama alarak bir araya getiren yaklaşım, değişken miktarının gözlem miktarından fazla olduğu durumlarda çok iyi performans sergilemektedir. Rastgele Orman büyük ölçekli veri setlerine uygulanabilmektedir ve değişkenlerin önemiyle ilgili bilgi verebilmektedir (Biau ve Scornet, 2016: 197).

Rastgele Orman, birden fazla Karar Ağacı'nın bir araya gelmesiyle oluşturulur. Daha sonra, her bir ağaç tek başına bir tahmin yapar ve sonuçlar birleştirilerek final sonuç elde edilir. Şekil 2.7'de Karar Ağacı ve Rastgele Orman Algoritması yapıları gösterilmektedir.



Şekil 2.7 Karar Ağacı ve Rastgele Orman Algoritması Yapıları

Kaynak: (Silipo, 2019)

Rastgele Orman, Karar Ağacı'na kıyasla daha yüksek bir doğruluk oranına sahip olabilir. Çünkü Karar Ağacı aşırı öğrenmeye yatkındır ve sadece eğitim verilerine odaklanırken,

Rastgele Orman, birçok ağacın tahminlerini bir araya getirerek daha genelleştirilmiş bir sonuç üretir. Bununla birlikte, Rastgele Orman, Karar Ağacı'na göre daha fazla hesaplama gücü gerektirir ve ayrıca modeli anlamak ve yorumlamak daha zordur (Ali vd., 2012: 274).

Karar Ağaçları Topluluk Öğrenmesi yöntemi ağaçların bağımsız olarak oluşturulması ve sonuçların birleştirilmesi nedeniyle Rastgele Orman yöntemine benzer. Ancak Rastgele Orman yönteminde ağaçlar rastgele özellik alt kümesi kullanarak büyütülürken, Karar Ağaçları Topluluk Öğrenmesi yönteminde farklı özellikler kullanılarak birçok farklı ağaç oluşturulabilir (Banfield vd., 2006: 173).

2.5.5. Lineer/ Doğrusal Regresyon (Linear Regression)

Veri kümesindeki bağımlı değişkenin (hedef değişken) bağımsız değişkenler (özellikler) tarafından açıklanmasını sağlayan bir modeldir. Bu model veri kümesindeki değişkenler arasındaki ilişkiyi doğrusal olarak ifade eder. Lineer Regresyon veri analizi sürecini öğrenmek için kullanılabilir.

Lineer regresyon, bir çıktı değişkeni ile bir ya da birden çok girdi değişkeni arasındaki ilişkiyi modeller. Özellikle tek girdi değişkeni olan lineer regresyon "tek değişkenli" veya "basit" lineer regresyon olarak adlandırılırken, birden fazla girdi değişkeni olan lineer regresyon "çok değişkenli" veya "çoklu" lineer regresyon olarak adlandırılır (Rong ve Bao-Wen, 2018: 2).

Lineer regresyonun avantajı, anlaşılmasının kolay olması ve düzenli hale getirme yoluyla aşırı öğrenmenin önlenmesinin de kolay olmasıdır (Ray, 2019: 36).

Ridge, LASSO (Least Absolute Shrinkage and Selection Operator) ve Elastic Net, lineer regresyon modellerinin aşırı öğrenme (overfitting) sorununu çözmek için kullanılan düzenlileştirme teknikleridir (Bonaccorso, 2017: 80).

Ridge hata terimine kareler toplamının bir çarpanı olan bir ceza terimi ekleyerek çalışır. Bu ceza terimi, modelin katsayılarının büyüklüklerinin artmasını engeller ve böylece modelin aşırı öğrenme (overfitting) yapmasını önler. Ridge regresyonu, özellikle yüksek boyutlu veri kümeleriyle çalışırken etkilidir (Alay, 2019: 1).

Lasso hem aşırı öğrenmeyi azaltmak için büyük regresyon katsayılarını küçülterek tahmin hatasını iyileştirebilir hem de büyük varyanstan sorumlu değişkenleri atarak değişken seçimi gerçekleştirerek modeli daha yorumlanabilir hale getirir (Papadaki ve Tsagris, 2020: 6). Lasso, Ridge'e benzer şekilde çalışır ancak kareler toplamı yerine katsayıların mutlak değerleri toplamının bir çarpanı olan bir ceza terimi ekler. Lasso Ridge'den farklı olarak veri

kümesindeki bazı özelliklerin tamamen atılmasına yol açar. Bu nedenle özellik seçimi için sıklıkla kullanılır (Alay, 2019: 1).

Elastic Net, Ridge ve Lasso'nun birleşimidir (García-Nieto vd., 2021: 17136). Bu yöntem Ridge ve Lasso'nun avantajlarını birleştirir ve dezavantajlarını aşabilmektedir. Elastic Net, Ridge ve Lasso'dan daha esnek bir yöntemdir ve özellikle düşük sinyal-gürültü oranına sahip veri kümelerinde etkilidir (Alay, 2019: 1).

2.5.6. XGBoost Learning

XGBoost (eXtreme Gradient Boosting), ağaç tabanlı bir öğrenme yöntemi olan Gradient Boosting'in bir türüdür. Gradient Boosting yöntemi birden fazla zayıf tahmin ediciyi bir araya getirerek güçlü bir tahmin edici oluşturur. Bu yöntemde her bir tahmin edici, önceki tahmincilerin hatalarını en aza indirmeye çalışarak eğitilir (Chen ve Guestrin, 2016: 785).

XGBoost, ağaçların sırayla oluşturulduğu bir topluluk öğrenme yöntemidir. Sıralama devam ettikçe hatalar azalır. Her ağaç bir önceki ağaçtan öğrenir ve kalan hatayı yeniler. En yeni ağaç bu şekilde en az artık hataya sahip olacaktır (Priyadarshini vd., 2022: 4)

XGBoost, Gradient Boosting yöntemine benzer şekilde çalışır ancak diğer yöntemlere kıyasla daha hızlı ve daha verimlidir (Can, 2020: 17). Bu yöntemde, her bir tahmin edici ayrı ayrı eğitilir ve ardından bir araya getirilerek final tahmin yapılır. Ayrıca XGBoost, ağaçların büyüklüğü, ağaç sayısı, öznitelik seçimi ve öğrenme oranı gibi hiperparametrelerin otomatik olarak ayarlanmasına olanak tanıyan birçok optimize edici özellik içerir.

XGBoost, sınıflandırma ve regresyon problemlerinde kullanılabilir ve yüksek doğruluk oranlarına sahiptir. Ayrıca, büyük boyutlu veri kümelerinde iyi performans gösterir ve hızlı öğrenme süreleri sunar. Bu nedenle son yıllarda XGBoost makine öğrenmesi topluluğu arasında çok popüler bir yöntem haline gelmiştir (Ghojogh ve Crowley, 2019: 211).

2.6. Değerlendirme Ölçütleri

Modellerde gerçek değerlerle tahmin edilen değerlerin arasındaki mesafeyi bilmek önemlidir. Regresyon modellerini değerlendirebilmek için kullanılan metrikler R Kare, Düzeltilmiş R Kare, Ortalama Kare Hatası (Mean Squared Error, MSE), Kök Ortalama Kare Hatası (Root Mean Squared Error, RMSE) ve Ortalama Mutlak Hata (Mean Absolute Error, MAE)'dır. Sınıflandırma modellerini değerlendirmek için ise Confusion Matrix (Hata Matrisi), Accuracy (Doğruluk), Precision (Kesinlik), Recall (Duyarlılık), F1 Skoru, ROC Eğrisi ve ROC-AUC Skoru metrikleri kullanılmaktadır (Luque vd., 2019: 218).

2.6.1. R Kare ve Düzeltilmiş R Kare (R-Squared ve Adjusted R-Squared)

R Kare (R-squared) bir istatistiksel ölçüdür ve bağımsız değişkenlerin bağımlı(hedef) değişkendeki varyansı açıklama yeteneğini ifade eder. R^2 değeri genellikle 0 ile 1 arasında değişir ve yüzde olarak yorumlanır. 1'e ne kadar yakınsa, bağımsız değişkenlerin bağımlı(hedef) değişkendeki varyansı açıklama yeteneği o kadar yüksek ve model hassas demektir (Barut ve Bilgin, 2023: 155). Daha yüksek R^2 değeri, bağımsız değişkenin bağımlı(hedef) değişken üzerindeki etkisinin daha yüksek olduğunu gösterir. Denklem 2.1'de R^2 'nin hesaplanma formülü bulunmaktadır (Barut ve Bilgin, 2023).

$$R^2 = \frac{\sum(x-\bar{x})(y-\bar{y})}{\sum(x-\bar{x})^2(y-\bar{y})^2} \qquad (2.1)$$

Özellikle çoklu regresyon analizinde birden fazla bağımsız değişken kullanıldığında, R Kare'nin yüksek olması bağımsız değişkenler arasındaki etkileşimleri göz ardı edebilir. Bu nedenle Düzeltilmiş R^2 (Adjusted R^2) kullanılarak R^2 değeri düzeltilebilir.

Düzeltilmiş R^2, birden fazla bağımsız değişkenin kullanıldığı regresyon analizlerinde R^2'nin etkileşimleri hesaba katması için düzenlenmiş bir istatistik değeridir. Düzeltilmiş R^2, birden fazla bağımsız değişkenin eklenmesiyle R^2'nin yükselebileceği gerçeğini de dikkate alır ve bu nedenle modelin gerçekçi bir değerlendirmesi için daha uygun bir ölçüttür.

2.6.2. Ortalama Kare Hatası (Mean Squared Error, MSE) ve Kök Ortalama Kare Hatası (Root Mean Squared Error, RMSE)

Ortalama Kare Hatası (Mean Squared Error, MSE), bir regresyon modelinin tahminlerinin gerçek değerlerden ne kadar uzak olduğunu ölçmek için kullanılan bir metriktir. MSE tahmin edilen değerler ile gerçek değerler arasındaki farkların karelerinin ortalamasını hesaplar. Daha düşük MSE değeri daha iyi bir model performansını gösterir. Ancak MSE'nin orijinal veri birimleri üzerinde kareler alınarak hesaplanması nedeniyle yorumu bazen zor olabilmektedir. Bu nedenle RMSE (Kök Ortalama Kare Hata) gibi daha yorumlanabilir alternatifler de tercih edilmektedir (Hodson, 2022: 5481-5487).

Kök Ortalama Kare Hatası (Root Mean Squared Error, RMSE), MSE'nin karekökünü alarak elde edilen bir regresyon performans metriğidir. RMSE modelin gerçek değerlerden ne kadar uzak olduğunu ölçer ve hata değerlerini orijinal veri birimine dönüştürür. RMSE gerçek değerle ve tahmin edilen değerlerin arasındaki farkın standart sapmasıdır (Özen vd., 2021: 137). RMSE değerinin genellikle daha küçük olması istenmektedir. Daha düşük RMSE değeri modelin daha iyi tahminler yaptığını ve gerçek değerlere daha yakın olduğunu göstermektedir.

2.6.3. Ortalama Mutlak Hata (Mean Absolute Error, MAE)

Ortalama Mutlak Hata (Mean Absolute Error, MAE) bir regresyon modelinin tahminlerinin gerçek değerlerden ne miktarda uzak olduğunu ölçmek için kullanılan bir metriktir. MAE tahmin edilen değerler ile gerçek değerler arasındaki mutlak farkların ortalamasını hesaplar (Özen vd., 2021: 136). Hataların büyüklüğünü ölçerken kare almak yerine doğrudan mutlak farkları kullanır. Bu sayede hata büyüklüklerinin daha dengeli bir şekilde değerlendirilmesini sağlar. Daha düşük bir MAE değeri modelin daha iyi tahminler yaptığını ve gerçek değerlere daha yakın olduğunu gösterir.

ÜÇÜNCÜ BÖLÜM
NBA'DE MAKİNE ÖĞRENMESİ UYGULAMALARI İLE İLGİLİ YAPILAN ÇALIŞMALAR

Alan yazında NBA'le ilgili yapılan çalışmalar incelendiğinde geçmiş yıllarda ekonometrik yöntemlerle ve iktisadi ölçütlerle uygulanmıştır. Literatürde bunlarla ilgili yapılmış bazı çalışmalara yer verilmiştir.

NBA için oyuncuların maaş tahmini, başarılı basketbol şutları ile ilgili sınıflandırma, oyuncularının maçlarda yaptığı atışların isabeti üzerine tahminleme, maç sonuçlarını tahminleme, final four oynama başarısını modelleme, basketbol antrenmanını gerçek zamanlı olarak analiz edebilen bir algoritma geliştirme, oyuncuların yaşadığı sakatlıklarla yaklaşan maçtaki riskin belirsizliğini azaltmayı amaçlayan, oyuncularının All-Star olmalarını tahminleme, oyuncularının performansında meydana gelen değişiklikleri inceleme, takımlar düzeyinde kazanma olasılığı ile oyun sonuçları arasındaki ilişkiyi inceleme, yükselen yıldız tahminini, All-Star Tahmini, Playoff Tahmini, oyuncularının All-Star maçına seçilip seçilmeyeceklerini tahminleme, oyun sonuçlarında en büyük etkiye sahip olan temel değişkenleri belirleme gibi çalışmalar literatürde yapılmıştır.

NBA'de maaş tahminiyle ilgili yapılan çalışmaları ve makine öğrenmesi uygulamalarını incelemek için kategoriler belirlenmiştir. NBA'de maaş tahmini ve makine öğrenmesi uygulamaları, tez konusunun ana unsurlarıdır. Bu nedenle literatürde bu konulara odaklanan çalışmaları incelemek, tezin kapsamını ve amacını doğru bir şekilde yansıtmak için önemlidir. Mevcut literatürde NBA'de maaş tahmini ve makine öğrenmesi uygulamalarıyla ilgili yapılan çalışmalar bulunmaktadır. Bu çalışmalar tez konusunun akademik bir ilgi alanı olduğunu ve mevcut bilgi birikimine dayandığını göstermektedir. Önceki çalışmaların eksik kaldığı ve yeterince keşfedilmemiş bu alanı ele alarak, teze yeni bir katkı yapmak istenmiştir. Bu nedenle literatürdeki araştırma boşluklarını belirlenmiş ve daha önceki çalışmaların eksik kaldığı alanlara odaklanılmıştır.

Bu tezdeki literatür taraması yapılırken NBA'de Maaş Tahminiyle İlgili Yapılan Çalışmalar, NBA'de Makine Öğrenmesi Uygulamaları ile İlgili Yapılan Çalışmalar, NBA'de Makine Öğrenmesi Uygulamaları Kullanılarak Maaş Tahminiyle İlgili Yapılan Çalışmalar başlıklarıyla araştırılmıştır.

3.1. NBA'de Maaş Tahminiyle İlgili Yapılan İstatiksel Çalışmalar

NBA oyuncularının maaşlarını tahmin etmek hem takımlar hem de oyuncular açısından önemli bir konudur. Takımlar, maaş bütçelerini etkili bir şekilde yönetmek için oyuncuların değerini doğru bir şekilde değerlendirmelidirler. Doğru bir maaş dağılımı, takımın başarısını arttırırken aynı zamanda finansal kaynakların etkin bir şekilde kullanılmasını sağlar. Ayrıca, oyuncuların maaşlarını tahmin etmek, takımların hangi oyuncularla uzun vadeli anlaşmalar yapacağına karar vermelerine yardımcı olabilir.

Diğer yandan, oyuncular açısından maaş tahmini, oyuncuların kariyerlerini planlamalarına yardımcı olacaktır. Maaşların tahmini, oyuncuların piyasa değerlerini anlamalarına ve gelecekteki sözleşmeleri için de planlamalarını yapmalarına yardımcı olabilir. Ayrıca oyuncuların diğer oyuncularla karşılaştırılarak nerede durduklarını öğrenilmesinde de rol oynar.

Literatür incelendiğinde yakın zamanda NBA için maaş tahminini istatiksel olarak yapan birçok çalışma bulunmaktadır.

NBA oyuncularının performansları ve oyuncuların maaşları arasındaki ilişkiyi Sigler ve Sackley (2000: 46-51) çalışmalarında araştırmışlardır. Ribaunt, asist ve maç başına sayı olmak üzere üç tahmin değişkeni ile maaş tahminini incelemişlerdir. Lineer Regresyonmodeli ve düşük bir belirleme katsayısı (R^2) ile çalışmasını hesaplamıştır. 1997-1998 sezonundan verileri kullanarak çalışmalarını yapmışlardır.

Ertug ve Castellucci (2013: 407-431) 1989-1990 ve 2004-2005 sezonu arasındaki yılların arasındaki verileri kullanarak bir çalışma gerçekleştirmişlerdir. Takım ürün kalitesi, takım gelirleri, yüksek itibarlı/yüksek statülü kaynak sağlayıcılar için tercih ve oyuncu maaşı değişkenlerini bağımlı değişken olarak kullanmışlardır. Kullandıkları lineer modellerinin R^2 değerleri 0.30 ve 0.31 olarak tespit edilmiştir. Araştırmacılar oyuncuların maaşlarını çoğu oyuncuların sahadaki performansıyla ilgili olmayan bir dizi tahmin değişkeni ile ilişkilendirmiştir.

Stanek (2016) yüksek lisans tezinde NBA oyuncusu maaş analizi yapmak için oyuncu performansını ve takım gelirlerini araştırmıştır. Bir maç galibiyetinin bir takım için değerini geliri açısından analiz eden Stanek ürün marjinal gelirine göre ligdeki oyuncuların ve süper yıldızların olası eksik veya fazla ödemelerini incelemiştir. Oyunculara, özellikle de süperstarlara fazla ödeme yapıldığını gösteren bulgulara ulaşmıştır.

Xiong vd. (2017: 55-61) NBA'de oyuncu maaşını tahminleme çalışması yapmıştır. 2013-2014 sezonu verileri kullanılan çalışmada oyuncuların sahadaki performansını ölçen çeşitli değişkenler kullanılmıştır. 4 farklı separate Multiple Lineer Regresyon analizi yaparak

bir çalışma gerçekleştirilmiştir. Multiple Lineer Regresyonda yedi tür oyuncu performansı ortalama maaşı tahmin etmek için test etmişlerdir. Sadece maç başına sayı, maç başına ribaund ve maç başına asist değişkenleri ortalama maaşla pozitif korelasyona sahiptir ve maç başına top çalma değişkeni ortalama maaşla negatif korelasyona sahip çıkmıştır. Oyuncu pozisyonu ve ortalama maaşlar arasında istatistiksel olarak anlamlı herhangi bir değişiklik olup olmadığını belirlemek için bir Anova testi yapmışlardır. Bu testi herhangi bir pozisyonun diğer pozisyonlardan daha fazla veya daha az para kazandırdığını görmek amacıyla gerçekleştirmişlerdir. Uyguladıkları ANOVA ve Tukey testinde her bir pozisyonun maaş ortalamaları arasında anlamlı bir fark var bulmuşlardır ama sadece Pivot ve Şutörlerin ortalama maaşları arasında bu fark ortaya çıkmıştır.

Maaş, iş performansı ve özgeci davranış arasındaki ilişkiyi incelemek için Asghar ve Asif (2018: 71-79) bir çalışma yapmıştır. 2012-2013 sezonu ile 2015-2016 sezonu arasındaki verileri kullanarak regresyon analizi uygulamışlardır. Maaşın özgeci davranışla ve iş performansıyla pozitif olarak ilişkili olduğunu tespit etmişlerdir. Maaşın bağımsız değişken olduğu regresyon analizinde regresyon katsayısı özgeci davranış için $\beta=0.665$, iş performansı için $\beta=0.846$ olarak çıkmıştır. Özgeci davranışın iş performansı üzerinde anlamlı pozitif etkisi olduğu ve maaş ile iş performansı arasındaki ilişkide aracılık etkisi gösterdiğini söylemektedirler. Çalışmada bahsedilen özgeci davranış bireyin başkalarının refahını veya çıkarlarını kendi çıkarlarının üzerinde tutma eğilimini ifade etmektedir (Karahalilöz, 2021: 53).

Sigler ve Compton (2018: 1-12) NBA oyuncularının maaşını belirlemede neyin önemli olduğunu analiz ederken 3 sayılık atış ve Hollinger oyuncu verimlilik derecesini (Player Efficiency Rating, PER) de dahil ederek oyundaki değişimi açıklayan bir çalışma yapmışlardır. Çalışmalarında 2017-2018 sezonu verilerini kullanmışlardır. Çalışmalarında modele bağımsız değişkenlerden her seferinde en az anlamlı olan bir tanesini eleyerek geriye doğru regresyon gerçekleştirmişlerdir. Puanların, oyuncu deneyiminin, asistlerin, ribauntların ve faullerin; maaş konusunda istatistiksel olarak önemli faktörler olduğunu ancak aynı zamanda yapılan 3 sayılık atışların ve Hollinger'in PER'ini önemsiz olarak bulmuşlardır.

Hentila (2019: 1-31) “The link between salary and performance: Are NBA players overpaid?” isimli tezinde oyuncuların performanslarıyla maaşları arasındaki ilişkiyi regresyon yöntemleriyle incelemiştir. 2017-2018 sezonunun verilerini kullanan Hentila oyuncunun yaşı, oynama pozisyonu ve takım sabit etkileri için oyuncu maaşı ile performans kontrolü arasında önemli bir pozitif ilişki olduğunu göstermiştir. Çalışmasında oyuncunun oyunlardaki toplam istatistiklere karşı genel istatistiksel katkısını ölçen PIE (Performance Impact Estimate)

ölçümünü kullanmıştır. Bazı oyuncuların performanslarına göre sınıfının altında bir maaş aldığını tespit etmiştir.

Çalışmalarında Kekez vd. (2021: 227-248) NBA'de bireysel sonuçlara, fiziksel özelliklere ve profesyonel sözleşme maaşlarına dayalı oyuncu seçimi ile ilgili kararları etkileyen faktörleri bulmak için veri madenciliği yöntemlerini kullanmıştır. 1946-2017 verilerini Rapidminer programıyla Korelasyon analizi, Lineer regresyon, Kümeleme yöntemleriyle tahminlemeye çalışmışlardır. Yüksek maaş sözleşmeleri olan oyuncuların birçoğu yüksek performans standartlarını karşılamadığından, bu oyuncular arasındaki tutarsızlıkları belirlemek için bir yapay sinir ağı algoritması kullanılmıştır. Oyuncuların verimlilik katsayıları ile takımların sonuçlarına katkıları arasında pozitif bir ilişki çıkmıştır. Ancak bu ilişkinin kuvveti farklılık göstermektedir. Oyuncuların fiziksel özellikleri ile maaşları arasında anlamlı bir ilişki bulunmamıştır. Bununla birlikte sonuçlara katkıda bulunduğu düşünülen özelliklerle zayıf pozitif ilişkiler gözlemlenmiştir. Oyuncu Verimlilik Derecesi olan PER ve Oyuncuya Göre Üstünlük Değeri olan VORP, lineer regresyon modelindeki iki ana faktör olarak kanıtlanmıştır. PER katsayısının maaşlar üzerindeki etkisi zamanla azalırken, Yedek Oyuncuya Göre Üstünlük Değeri (Value Over Replacement Player, VORP) katsayısının etkisi artmakta olduğunu söylemektedirler. Araştırmacılar VORP değişkeninin maaşlarla güçlü bir şekilde ilişkili olduğunu bulmuşlardır.

Papadaki ve Tsagris'in (2020: 1-19) çalışmasında dikkat çektiği gibi bu tür analizlerden elde edilen sonuçların yeterince geçerli ve güvenilir olup olmadığı sorusu literatüre başka bir bakış açısıyla bakmayı gerektirmiştir. Son yıllarda makine öğrenmesi teknikleriyle maaş tahminleme yapmak popülerlik kazanmıştır.

3.2. NBA'de Makine Öğrenmesi Uygulamaları ile İlgili Yapılan Çalışmalar

Literatürde NBA verileriyle en çok oyun sonucu tahmini, oyuncu performans tahminleri yapıldığı görülmüştür. Oyuncu maaş tahminlerine literatürde çok az rastlanmıştır.

Türkiye'de NBA verileriyle oyun sonuçları ve oyunlarla ilgili tahminlemelerin makine öğrenmesi yöntemleriyle yapan tez çalışmaları bulunmaktadır. Ancak maaş tahminlemelerine ilişkin çalışma bulunmamaktadır. Bu bölümde NBA, Euroleague ve Türkiye Erkekler Basketbol Süper Ligi oyun sonuçlarıyla ilgili makine öğrenmesi yöntemleri kullanan tez çalışmalarına yer verilmiştir.

Yığman (2018: 1-43) tez çalışmasında Euroleague verileriyle makine öğrenmesi kullanarak başarılı basketbol şutları ile ilgili sınıflandırma tahmin modelleri geliştirmiştir. Lojistik Regresyon %60.14, Lojistik Regresyon Interaction Term %60.17, Rastgele Orman

%60.72, Naive Bayes %60.2, SVM %60.16, Yapay Sinir Ağları %60.2 modellerini doğruluk oranlarıyla bulmuştur. Rastgele Orman algoritmasının en yüksek doğruluk oranıyla diğer modeller arasından en başarılı algoritma olduğunu belirtmiştir.

NBA oyuncularının maçlarda yaptığı atışların isabeti üzerine Kılıç Depren (2019: 256-269) makine öğrenmesi yöntemiyle bir çalışma yapmıştır. 2014-2015 sezon verilerini kullanan araştırmacı dokuz bağımsız değişken ve bir ikili bağımlı değişkenle çalışmıştır. Yedi makine öğrenmesi yöntemi kullanan Kılıç Deprenin algoritmaların doğruluk oranları Lojistik Regresyon için %48.6, Linear Discriminant Analysis için %48.6, C5.0 için %37.3, K-En Yakın Komşu Algoritması için %57.5, Naive Bayes için %58.4, Çok Katmanlı Algılayıcılar için %27.2 ve Multivariate Adaptive Regression Splines için %41.8 olarak tespit edilmiştir.

Özdemir (2020: 1-56) yüksek lisans tezinde Türkiye Erkekler Basketbol Süper Ligi maç sonuçlarını tahminleyen modeller geliştirmiştir. K-En Yakın Komşu Algoritması, Lojistik Regresyon, Çok Katmanlı Sinir Ağı, Naive Bayes, J48 ve Oylama yöntemleri kullanılmıştır. Farklı veri setleri kullanarak birçok modelle tahminleme yaparak farklı sonuçlar elde etmiştir. Özdemir E. Tüm veri setleri üzerinde yaptığı çalışmalar sonucunda Lojistik Regresyon ve Çok Katmanlı Algılayıcı algoritmalarının diğer modellerden daha başarılı olduğunu belirtmiştir.

Kurum (2020: 1-71) yüksek lisans tezinde NBA yarışmasının sonucunu tahmin etmek için makine öğrenmesi yöntemlerini kullanmıştır. Naive Bayes algoritması, Destek Vektör Makinesi, K-En Yakın Komşu Algoritması, Sinir Ağları, Rastgele Orman ve Karar ağacı algoritmalarını kullanan Kurum maç kazananını tahmin etmek için en başarılı algoritmanın Destek Vektör Makine algoritması olduğunu tespit etmiştir. Karar Ağacı, Rastgele Orman ve Yapay Sinir Ağı Algoritmasının doğruluk oranları sırasıyla %63, %63, %58 olarak bulunmuştur. En başarısız algoritmalar olarak Naive Bayes %51 ve K-En Yakın Komşu Algoritması %52 doğruluk oranıyla olduğunu belirtmiştir. %67 doğrulukla en iyi algoritmanın Destek Vektör Makinesi olduğunu söyleyen Kurum tezinin ikinci aşamasında basketbol maçı başına düşen puan sayısı tahminlemeyi Lineer Regresyon ve Genelleştirilmiş Doğrusal Modelleri kullanarak Kök ortalama kare hatası sırasıyla 20.5 ve 21.8 bulmuştur.

Erdoğan (2021: 1-72) Euroleague'de oynanan maçların sonucunun ve final four oynama başarısının makine öğrenmesi ile tahmin edilmesi üzerine tez çalışması yapmıştır. Lojistik regresyon, Yapay Sinir Ağları, Rastgele Orman, Naive Bayes ve Karar Ağacı yöntemlerini sınıflandırma modellemesi için kullanmıştır. Maç sonucu tahminlerinde kullanılan modellerin doğruluk oranlarını Lojistik Regresyon %91.78, Yapay Sinir Ağları %91.48, Rastgele Orman %90.13, Naive Bayes %88.77, Karar Ağacı %82.51 olarak belirtmiştir. Final four oynama

başarısını tahminlemek için Yapay Sinir Ağları v e Lojistik Regresyon kullanmıştır ve bu modellerin doğruluk oranları sırasıyla %77.02 ve %76.9 olarak tespit etmiştir.

Teket (2021: 1-40) yüksek lisans tezinde bir basketbol antrenmanını gerçek zamanlı olarak analiz edebilen bir algoritma geliştirmiştir. YOLOv3-tiny kullanarak bir video analizi yapmıştır. NBA oyununda bu algoritmayı denemiş ve %80 üstünde doğruluk tespit etmiştir. Fakat gerçek zamanlı kullanımının yüksek zaman engeli yüzünden mümkün olmadığı sonucuna ulaşmıştır.

Kahraman (2022: 1-81) yüksek lisans tez çalışmasında NBA ve TBL'de oyunların sonuç tahminini makine öğrenmesi yöntemleriyle geliştirmeyi hedeflemiştir. Çalışmasında ELO Rating Modeli, Lojistik Regresyon, Destek Vektör Sınıflandırıcısı, Karar Ağacı, Rastgele Orman, Naive Bayes, KNN, LGBM, XGBoost ve Yapay Sinir Ağları modelleri kullanmıştır. Doğruluk oranları NBA için ELO Rating Modeli %66.2, Lojistik Regresyon %67.5, Destek Vektör Sınıflandırıcısı %67, Karar Ağacı %63.3, Rastgele Orman %66.2, Naive Bayes %65.6, KNN %65.1, LGBM %65.1, XGBoost %65.6 ve Yapay Sinir Ağları %64 olarak tespit edilmiştir. Lojistik Regresyon ve Naive Bayes algoritmalarının en başarılı tahminleme yapan algoritmalar olduğu sonucuna ulaşmıştır.

Diğer Çalışmalar

Makine öğrenmesi teknikleri kullanılarak Basketbol liglerinde tahminleme çalışması yapan çalışmalara 2015 sonrasında rastlanmaktadır.

Karaoğlu (2015: 1-5) spor karşılaşmasının sonucunu tahmin eden bir makine öğrenmesi modeli geliştirmiştir. Takımların ofansif ve defansif yetenek değişkenlerinin maç sonucunu tahminlemeye çalışılmıştır. Avrupa'nın on altı futbol liginden veriler toplanarak uygulanmıştır. Naive Bayes, BayesNet, Çok Katmanlı Algılayıcılar, LogitBoost, Decision Table, ZeroR, C4.5 algoritmalarını kullanan araştırmacılar liglerde en yüksek tahminlemeyi en çok Decision Table algoritmasıyla yakalamışlardır.

NBA maç sonuçlarını makine öğrenmesi yöntemleriyle tahminleme çalışması yapan Ayyıldız (2018: 40-53) 2015-2016 sezonunun verilerini kullanmıştır. Kazandı ve kaybetti olarak iki değerli çıktısı olacağı için sınıflandırma algoritmalarında da kullanılan yapay sinir ağlarını kullanmıştır. Yapay sinir ağlarını 2 katmanlı oluşturmuş ve eğitim de TRAINLM (Levenberg-Marquardt backpropagation) algoritmasını kullanmıştır. Oluşturduğu modelin doğruluk değeri %89.5 olarak bulunmuştur.

Thabtah vd. (2019: 103-116) NBA'da oynanan oyunların sonuçlarını tahmin etmeyi ve oyun sonuçlarını etkileyen önemli faktörleri belirlemeyi amaçlamışlardır. Naive Bayes, Yapay Sinir Ağı ve Karar Ağacı algoritmalarını kullanmışlardır. Analizlere göre savunma ribauntları

DRB (defensive rebounds, DRB) oyununun sonucunu etkileyen en önemli değişken olarak bulmuşlardır. Üçlük yüzdesi (TPP), serbest atışlar (FT) ve toplam ribauntlar (TRB) gibi önemli faktörleri eklendiğinde modelin tahminleme doğruluk oranını %2 ve %4 arasında arttığını tespit etmişlerdir.

Özdemir ve Ballı (2020: 740-752) çalışmasında 2017-2018 sezonunda Türkiye Erkekler Basketbol Süper Ligi maçları, Dört Faktör ve DefansOfans modeli ile birlikte makine öğrenmesi yöntemleri kullanarak analiz etmişlerdir. Makine öğrenmesi alanında popüler yöntemlerden olan kNN, Lojistik Regresyon, Çok Katmanlı Algılayıcı, Naive Bayes, j48 ve Oylama yöntemleri kullanılmıştır. Elde edilen sonuçlara göre Lojistik Regresyon en iyi tahmin başarısıyla %96,55 olarak belirlenmiştir.

NBA maçlarında oyuncuların yaşadığı sakatlıklarla ilgili Wu (2020: 295-308) 2017 verilerini kullanarak bir makine öğrenmesi çalışması yapmıştır. Yaklaşan maçtaki riskin belirsizliğini azaltmayı amaçlamışlardır. Oyuncuların temel istatistikleri, önceki maçlardaki performansları, rakip takımın durumu gibi verileri kullanmışlardır. Değişkenlerin önemini belirlemek için Rastgele Orman Sınıflandırma algoritması kullanmışlardır. Farklı özellikler arasındaki korelasyonu göstermek ve veri boyutunu azaltmaya yardımcı olmak için Temel Bileşenler Analizi (The Principal Component Analysis, PCA)'de uygulanmıştır. Sınırlı ve dengesiz veri setinden dolayı modelin sakatlık tahmininde sınırlı bir güce sahip olduğunu ancak takımdaki yıldız oyuncunun sakatlığı konusunda iyi bir sonuç elde ettiğini söylemektedirler. Oyuncunun kendi performansının sakatlık konusunda en önemli faktör olduğunu tespit etmişlerdir.

Albert vd. (2021: 97) çalışmasında hibrid bir makine öğrenmesi teknikleri kullanarak NBA oyuncularının All-Star olmalarını tahminleme çalışması gerçekleştirmiştir. Rastgele Orman % 87.3 doğruluk oranıyla, Adaboost % 85.2 doğruluk oranıyla, MLP % 88.7 doğruluk oranıyla tahminleme yapmıştır. Ardından hibrit bir yaklaşım sunmak amacıyla yapay sinir ağları algoritması kullanan araştırmacılar % 88.7 doğruluk oranı elde etmişlerdir.

Georgievski ve Vrtagic (2021: 3339-3343) NBA oyuncularının performansında meydana gelen değişiklikleri veri simülasyonu yoluyla analiz etmeyi amaçlamışlardır. Makine öğrenmesi yöntemlerinden Çok Katmanlı Algılayıcılar (MLP) algoritması kullanan araştırmacılar MAE değerini 0.072 olarak bulmuşlardır.

Liv vd. (2021: 102123) NBA takımları düzeyinde kazanma olasılığı ile oyun sonuçları arasındaki ilişkiyi incelemek için çok değişkenli lojistik regresyon analizi kullanmıştır. Ardından veri zarflama analizi (DEA) kullanarak bir oyuncu portföyü verimlilik analizi yaparak oyuncuları optimal olarak seçmeyi ve oyuncular arasında oyun süresini planlamayı

amaçlamıştır. 2011-2012 ve 2014-2015 NBA sezonları arasındaki verileri kullanmışlardır. Golden State Warriors takımının verilerini alarak Lojistik Regresyon algoritmasıyla bir makine öğrenmesi çalışması yapmışlardır. Algoritma %82.1 oranında doğruluk oranıyla tahminleme yapmıştır.

Mahmood vd. (2021: 106506) çalışmalarında basketbolda yükselen yıldız tahminini bir makine öğrenimi problemi olarak sunmuşlardır. Farklı veri setleriyle makine öğrenmesi modeli olarak Classification And Regression Trees (CART), Destek Vektör Makineleri, Maximum Entropy Markov Model, Bayesian Network, Naive Bayes tekniklerini kullanmışlardır. Veri setleri içindeki F ölçümlerine göre en yüksek oranda CART ve SVM algoritmaları %89 olarak tespit edilmiştir. %87 ölçümüyle Maximum Entropy Markov Model ve Naive Bayes algoritmaları ikinci sırada başarı göstermişlerdir.

NBA oyunlarındaki verilerle makine öğrenmesi yöntemleri kullanarak Wang ve Fan (2021: 032036) All-Star Tahmini, Playoff Tahmini ve Sıcak Seri Yanılgısı (Hot Streak Fallacy) tahminleme yapmışlardır. Araştırmacılar All-Star tahminleme K-En Yakın Komşu Algoritması uygulayarak %95.83 doğruluk oranı, Lojistik Regresyon uygulayarak %97.9 doğruluk ve Karar Ağacı uygulayarak %95-%94 aralığında doğruluk oranı elde ederek yapmışlardır. Playoff Tahmini için Rastgele Orman algoritmasını kullanmışlardır. Algoritma %85 doğruluk oranıyla başarılı tahminleme yapmıştır. Sıcak seri, bir oyuncunun veya takımın ardışık olarak iyi performans gösterdiği bir dönemi ifade eder. Bu yanılgıyı tahminletmek için Lineer Regresyon ve Basit Regresyon kullanılmıştır. İki modelde %40 doğruluk değerinin altında kalmıştır.

Çene (2022: 29-52) Euroleague'de oynanan maç sonuçlarını tahminlemek için bir çalışma yapmıştır. 2016-2017 ve 2020-2021 sezonları arasındaki verileri toplamıştır. Çalışmasında K-En Yakın Komşular, Naive Bayes, Lojistik Regresyon, SVM, Karar Ağaçları, Rastgele Orman ve Yapay Sinir Ağları algoritmaları kullanarak en iyi tahmin modelini geliştirmeyi amaçlamıştır. Bu algoritmalar arasından en iyi tahminleme yapanları Lojistik Regresyon %84.4, SVM %84.2 ve Yapay Sinir Ağları %84.5 doğruluk oranlarıyla tespit etmiştir. Savunma ribaundu, gerçek şut yüzdesi, top çalma, top kaybı, hücum ribaundu ve denenen serbest atış değişkenlerini sonuç tahmin modeli üzerinde en çok etki eden değişkenler olarak bulmuştur.

NBA oyuncularının All-Star maçına seçilip seçilmeyeceklerini tahminlemek için Nguyen vd. (2022: 217-235) makine öğrenmesi ve derin öğrenme çalışması yapmıştır. Hem sınıflandırma hem regresyon analizi yapan araştırmacı her yöntem için birden çok algoritma kullanmıştır. Regresyon analizi için Lineer/ Doğrusal Regresyon, Gradient Boosting Machine, Destek Vektör Makineleri, Rastgele Orman ve Yapay Sinir Ağları kullanmıştır. Regresyon

analizinin sonucunda PTS en önemli değişken olmak üzere en etkili 10 değişken tespit edilmiştir. Sınıflandırma analizi için Lojistik Regresyon, Stochastic Gradient Descent, Destek Vektör Makineleri, Rastgele Orman ve Naive Bayes kullanmıştır. Sınıflandırma algoritmasında da en etkili değişken PTS (Total Points) olarak bulunmuştur. Dengeli alt örnekleme yapısıyla oluşturulan Rastgele Orman modelinin, All-Star etkinliklerine seçilmek için yüksek potansiyele sahip NBA oyuncularını tespit etmek için en uygun algoritma olarak tespit etmişlerdir.

NBA oyun sonuçlarında en büyük etkiye sahip olan temel değişkenleri belirleme hedefiyle Patrod vd. (2023: 1-6) bir makine öğrenmesi modeli geliştirmiştir. Lineer Regresyon, SVM ve Karar Ağaçları algoritmalarını kullanmıştır. Sırasıyla algoritmaların doğruluk oranları %92, %85 ve %68 olarak bulunmuştur. En başarılı tahminlemeyi Lineer regresyon algoritması yapmıştır.

Horvat vd. (2023: 798) NBA ve diğer basketbol ligi maçlarının sonuçlarını tahmin etmek için makine öğrenimi yöntemlerini kullanarak bir model geliştirmişlerdir. Çalışmalarında matematiksel bir formülasyonla basketbol istatistiksel miktarları ve bunlardan türetilen performans göstergelerini açıklamışlardır. Lojistik Regresyon, Naive Bayes Karar Ağacı, Çok Katmanlı Algılayıcılar, K-En Yakın Komşular, Rastgele Orman, LogitBoost algoritmalarını kullanmışlardır. K-En Yakın Komşular %57.9 doğruluk oranıyla en iyi çalışan algoritma olmuştur. Bu yöntemlerle doğruluk oranlarını düşük bulan araştırmacılar yeni bir tahmin modeli geliştirmişlerdir. Yeni yöntemleriyle ortalama elde edilen tahmin doğruluğu yaklaşık %66, maksimum elde edilen doğruluk ise yaklaşık %78 olarak bulunmuştur.

3.3. NBA'de Makine Öğrenmesi Uygulamaları Kullanılarak Maaş Tahminiyle İlgili Yapılan Çalışmalar

Literatürde NBA oyuncularının maaşlarını makine öğrenmesiyle tahminleme ile ilgili son birkaç yılda çok az sayıda çalışmaya rastlanmıştır.

"Estimating NBA Players Salary Share According To Their Performance On Court: A Machine Learning Approach" isimli çalışmada Papadaki ve Tsagris (2020: 1-19) oyuncuların sahadaki performanslarına dayalı olarak NBA oyuncularının maaşını tahmin etmek için model geliştirmiştir. Çalışmada 2017-2019 yılları arasındaki oynanan NBA maçlarının istatistiklerinden; takım bordolarını, oyuncu maaşlarını ve oyuncu deneyimlerini birleştirerek 54 değişkenli bir veri seti toplanmıştır. Losso regresyonla değişken seçimi yapılmış ve Rastgele Orman algoritmasıyla tahminlemede bulunulmuştur. LASSO'nun ceza parametresini (λ) ayarlamak için 10 katlı çapraz doğrulama uygulanmıştır. Daha sonra Rastgele Orman algoritmasına eklenen en önemli faktörleri seçmek için seçilen λ değerini kullanarak LASSO

cezalandırma yöntemi kullanılmıştır. Analizin ikinci kısmında düşük ve yüksek ücretli oyuncular olarak iki sınıf arasında ayrım yapılmak istenmiş ve bunun için tekrar LASSO ve Rastgele Orman algoritmaları kullanılmıştır. 2017-2018 ve 2018-2019 tahminleri için AUC değerleri sırasıyla 0.811 ve 0.841 olarak tespit edilmiştir ve çapraz doğrulama sürecinin sonuçlarını doğruladığını belirtmişlerdir.

William vd. (2020: 1-21) NBA oyuncularının free agent(serbest oyuncu) dönemindeki maaşlarını incelemiştir. Bu oyunculara verilen maaş miktarlarını tahmin eden sınıflandırma tabanlı bir model geliştirmiştir. Korelasyon analizi yaparak puanların, top kayıplarının, yedek oyuncuya göre üstünlük değerinin (Value Over Replacement Player, VORP) ve ribauntların ağırlıksız KNN sınıflandırması için dengeli ve etkili bir değişken seti oluşturduğunu tespit etmişlerdir. Maaş olarak ortalama yıllık maaşa odaklanılmıştır ve sözleşmelerin tamamen garanti altına alınması nedeniyle NBA maaşlarının tipik arka uç ağırlığını göz ardı etmişlerdir. Model için 2014 ile 2017 yılları arasında imzalanan yaklaşık 500 sözleşme hakkında veri toplamışlardır. Oluşturacak değişkenleri seçmek için 2016-2017 sezonundaki oyuncu verilerini ve 2017 sezonundaki serbest oyuncu transferlerini kullanmışlardır. Standardizasyon ve 8'e ayırdıkları kümeleme sürecinden sonra, veri seti bir K-En Yakın Komşular sınıflandırma modelini eğitmek için kullanmışlardır. Dört değişken arasındaki ağırlıksız Öklid mesafesindeki 17 en yakın komşu kullanılarak test setinde sınıflandırma yapıldı. Modelin performansını ayarlamak ve değerlendirmek için en az mutlak hatalar olarak da bilinen bir L1-norm hata fonksiyonu kullanıldı. Test setinin ortalama hatası, 1.21'lik bir küme boyutuydu. Araştırmacıların bulgularına düşük ve orta düzey oyunculardaki hacim istatistiklerinin maaşlar için en belirleyici olduğu, yüksek VORP'un ise neredeyse maksimum maaş alan tüm oyuncular arasında yaygın değişken olduğu ortaya çıkmıştır.

Özbalta vd. (2022) NBA 2K20 MyTeam video oyunundan elde edilen veri seti ve NBA oyuncularının 2019-2020 sezonuna ait performans istatistikleri kullanarak makine öğrenimi yöntemleri ile NBA oyuncularının yeni sözleşmeleri için maaşları tahmin etme çalışması yapmışlardır. En iyi alt kümeleri seçmek için büzülme yöntemleri kullanan Özbalta vd. regresyon ve karar ağacı algoritmaları kullanarak tahminleme yapmışlardır.

Jain vd. (2022: 1-5) NBA'de oynayan bir basketbolcunun maaşını belirlemek için performansa ve diğer hususlara dayalı doğrusal olmayan bir yaklaşım önermişlerdir. Performans, yaş, pozisyon ve fantezi puanı değişkenlerini literatüre bakıldığında maaşı etkileyen çeşitli kategori olarak tespit etmişlerdir. Araştırmacılar 2020-2021 sezon verilerini kullanmışlardır. Oyuncu performanslarının maaşa etkisini incelemek için Pearson Korelasyon Katsayısı kullanılmıştır. Multiple Doğrusal Regresyon, Karar Ağacı, Rastgele Orman ve

Xgboost algoritmaları olmak üzere 4 regresyonla analiz yapılmıştır. Test veri seti ile modellerin doğruluğunu ölçmek için kök ortalama kare hataları (RMSE), ortalama mutlak hatalar (MAE) ve belirleme katsayısı (R^2) kullanılmıştır. R^2 oranları Multiple Lineer Regresyon için %57.04, Karar Ağacı için %64.4, Rastgele Orman için %75.85 ve Xgboost algoritması için % 88.18 olarak bulunmuştur.

Literatür taraması ile ulaşılan çıkarımlar sonucunda yapılması planlanan tez çalışmasının diğer çalışmalardan farkı, oyuncuların istatistiklerinin maaşa etkisinin birden çok değişkenle karşılaştırılarak çok sayıda makine öğrenmesi teknikleriyle analiz edilmesi olacaktır.

Oyuncuların maaş tahmini için hangi modellerin ve girdi setlerinin en uygun olduğunu anlamak için farklı makine öğrenmesi algoritmalarının karşılaştırılmalı ve farklı girdi değişkenleri kullanılmalıdır. En iyi maaş tahmini çalışmalarının çoğu yalnızca belirli bir modele odaklanır, dolayısıyla bu maaş tahmini modellemesi için hangi modellerin ve girdilerin önemli olduğunu göstermek açısından faydalı olabilir. Tablo 3.1'de Literatür taramasının tablolaştırılmış hali bulunmaktadır.

Tablo 3.1 Literatür taraması

	Çalışma	**Yıl**	**Sezon**	**Tahmin**	**Metot**	**Başarı**
1	Sigler ve Sackley	2000	1997-1998	Oyuncu Maaş Tahmini	Doğrusal Regresyon	Düşük bir belirleme katsayısı (R^2) elde etmişlerdir.
2	Ertug ve Castellucci	2013	1989-1990 ve 2004-2005 arası	Oyuncu Maaş Tahmini	Lineer Modellerinin	R^2değerleri 0.30 ve 0.31 olarak tespit edilmiştir.
3	Stanek	2016		Oyuncuların ve Süper Yıldızların Eksik veya Fazla Ödemeleri	Ürün Marjinal Geliri	Süperstarlara fazla ödeme yapıldığını gösteren bulgulara ulaşmıştır.
4	Xiong vd.	2017	2013-2014	Oyuncu Maaş Tahmini	Multiple Linear Regresyon Analizi, ANOVA ve Tukey Testi	Maç başına sayı, maç başına ribaund ve maç başına asist değişkenleri ortalama maaşla pozitif korelasyona sahiptir ve maç başına top çalma değişkeni ortalama maaşla negatif korelasyona sahip çıkmıştır.
5	Asghar ve Asif	2018	2012-2013 ve 2015-2016 arası	Oyuncu Maaş Tahmini, İş Performansı ve Özgeci Davranış Arasındaki İlişki	Regresyon Analizi	Maaşın bağımsız değişken olduğu regresyon analizinde regresyon katsayısı özgeci davranış için β=0.665, iş performansı için β=0.846 olarak çıkmıştır.
6	Sigler ve Compton	2018	2017-2018	Oyuncu Maaş Tahmini	Geriye Doğru Regresyon	Puanlar, oyuncu deneyimi, asist, ribaunt ve faullerin; maaş konusunda istatistiksel olarak önemli faktörler olduğunu ancak aynı zamanda yapılan 3 sayılık atışların ve Hollinger'in PER'ini önemsiz olarak bulmuşlardır.

7	Hentila	2019	2017-2018	Oyuncu Maaş Tahmini	PIE (Performance Impact Estimate) Ölçümü	Oyuncunun yaşı, oynama pozisyonu ve takım sabit etkileri için oyuncu maaşı ile performans kontrolü arasında önemli bir pozitif ilişki olduğunu göstermiştir. Bazı oyuncuların performanslarına göre sınıfının altında bir maaş aldığını tespit etmiştir.
8	Kekez vd.	2021	1946-2017	Bireysel Sonuçlara, Fiziksel Özelliklere ve Profesyonel Sözleşme Maaşlarına Dayalı Oyuncu Seçimi ile İlgili Kararları Etkileyen Faktörler	Korelasyon Analizi, Lineer Regresyon, Kümeleme	Oyuncuların verimlilik katsayıları ile takımların sonuçlarına katkıları arasında pozitif bir ilişki çıkmıştır. PER (Oyuncu Verimlilik Derecesi) katsayısının maaşlar üzerindeki etkisi zamanla azalmıştır, Yedek Oyuncuya Göre Üstünlük Değeri (Value Over Replacement Player, VORP) katsayısının etkisi artmıştır.
9	Yığman	2018		Başarılı Basketbol Şutları	Logistic Regression, Logistic Regression Interaction Term, Random Forest, Naive Bayes, Svm, Yapay Sinir Ağları	Logistic Regression %60.14, Logistic Regression Interaction Term %60.17, Random Forest %60.72, Naive Bayes %60.2, SVM %60.16, Yapay Sinir Ağları %60.2 doğruluk oranları elde edilmiştir.
10	Kılıç Depren	2019	2014-2015	Oyuncularının Maçlarda Yaptığı Atışların İsabeti	Logistic, Linear Discriminant Analysis C5.0, K-Nearest Neighbor, Naive Bayes, Multilayer Perceptron ve Multivariate Adaptive Regression Splines	Logistic Regression için %48.6, Linear Discriminant Analysis için %48.6, C5.0 için %37.3, k-Nearest Neighbor için %57.5, Naive Bayes için %58.4, Multilayer Perceptron için %27.2 ve Multivariate Adaptive Regression Splines için %41.8
11	Özdemir	2020		Maç Sonuçları	Knn, Lojistik Regresyon, Çok Katmanlı Sinir Ağı, Naive Bayes, J48 ve Oylama	Lojistik Regresyon ve Çok Katmanlı Algılayıcı algoritmalarının diğer modellerden daha başarılı olduğu belirtilmiştir
12	Kurum	2020		Maç Sonuçları	Naif Bayes, Destek Vektör Makinesi, K-NN, Sinir Ağları, Rastgele Orman ve Karar Ağacı	Destek Vektör Makinesi, Karar Ağacı, Rastgele Orman, Sinir Ağı, Naif Bayes ve k-NN algoritmasının doğruluk oranları sırasıyla %67, %63, %63, %58, %51, %52 doğruluk oranıyla olduğunu belirtmiştir.
13	Erdoğan	2021		Maçların Sonuçları ve Final Four Oynama Başarısı	Lojistik Regresyon, Yapay Sinir Ağları, Random	Doğruluk oranlarını maç sonuçları için Lojistik Regresyon %91.78, Yapay Sinir Ağları %91.48, Random Forest %90.13, Naive Bayes %88.77, Karar Ağacı

					Forest, Naive Bayes ve Karar Ağacı	%82.51 olarak belirtmiştir. Final four oynama başarısı için Yapay Sinir Ağları v e Lojistik Regresyon doğruluk oranları %77.02 ve %76.9 olarak tespit edilmiştir.
14	Teket	2021		Basketbol Antrenmanını Gerçek Zamanlı Olarak Analizi	Yolov3-Tiny Kullanarak Bir Video Analizi	%80 üstünde doğruluk tespit etmiştir. Gerçek zamanlı kullanımının yüksek zaman engeli yüzünden mümkün olmadığı sonucuna ulaşmıştır.
15	Kahraman	2022		Maç Sonuçları	ELO Rating Modeli, Lojistik Regresyon, Destek Vektör Sınıflandırıcısı, Karar Ağacı, Random Forest, Naive Bayes, KNN, LGBM, Xgboost ve Neural Network	Doğruluk oranları ELO Rating Modeli %66.2, Lojistik Regresyon %67.5, Destek Vektör Sınıflandırıcısı %67, Karar Ağacı %63.3, Random Forest %66.2, Naive Bayes %65.6, KNN %65.1, LGBM %65.1, XGBoost %65.6 ve Neural Network %64 olarak tespit edilmiştir.
16	Karaoğlu	2015		Maç Sonuçları	Naive Bayes, Bayesnet, Multilayer Perceptron, Logitboost, Decision Table, Zeror, C4.5	En yüksek tahminlemeyi en çok Decision Table algoritmasıyla yakalamışlardır.
17	Ayyıldız	2018	2015-2016	Maç Sonuçları	Yapay Sinir Ağları	Oluşturduğu modelin doğruluk değeri %89.5 olarak bulunmuştur.
18	Thabtah vd.	2019		Maç Sonuçları ve Oyun Sonuçlarını Etkileyen Önemli Faktörlerin Analizi	Naive Bayes, Yapay Sinir Ağı ve Karar Ağacı	Savunma ribauntları (defensive rebounds, DRB) oyun sonucunu etkileyen en önemli değişken olarak bulunmuştur. Üçlük yüzdesi (TPP), serbest atışlar (FT) ve toplam ribauntlar (TRB) gibi önemli faktörleri eklendiğinde modelin tahminleme doğruluk oranını %2 ve %4 arasında arttığını tespit etmişlerdir.
19	Özdemir ve Ballı	2020	2017-2018	Maç Sonuçları	Knn, Lojistik Regresyon, Çok Katmanlı Algılayıcı, Naive Bayes, J48 ve Oylama	Lojistik Regresyon en iyi tahmin başarısıyla %96,55 olarak belirlenmiştir.
20	Wu	2020	2017-2018	Oyuncuların Sakatlık Tahmini	Random Forest, Temel Bileşenler Analizi (The Principal Component Analysis, Pca)	Modelin sakatlık tahmininde sınırlı bir güce sahip olduğunu ancak takımdaki yıldız oyuncunun sakatlığı konusunda iyi bir sonuç elde ettiğini söylemektedirler. Oyuncunun kendi performansının en önemli faktör olduğunu tespit etmişlerdir.
21	Albert vd.	2021		All-Star	Random Forest, Adaboost, Mlp,	Random Forest % 87.3, Adaboost % 85.2, MLP % 88.7, Yapay Sinir

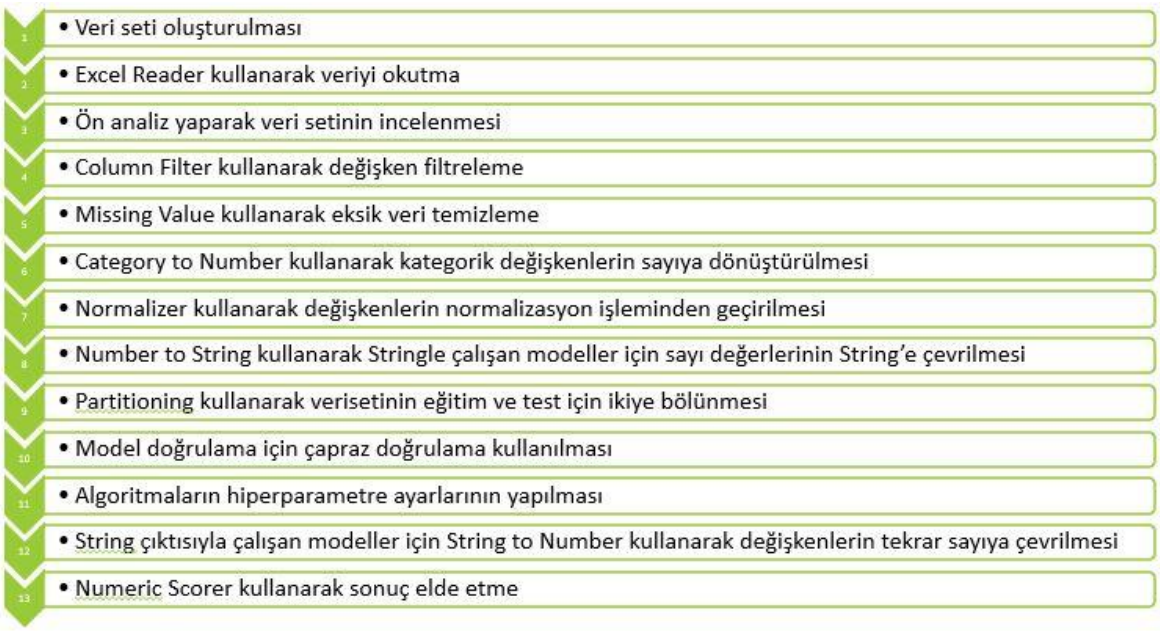
					Yapay Sinir Ağları	Ağları % 88.7 doğruluk oranıyla tahminleme yapmıştır.
22	Georgievski ve Vrtagic	2021		Oyuncularının Performansında Meydana Gelen Değişiklikler	Multilayer Perceptron (Mlp)	MAE değerini 0.072 olarak bulmuşlardır.
23	Liv vd.	2021	2011-2012 ve 2014-2015 arası	NBA Takımları Düzeyinde Kazanma Olasılığı ile Maç Sonuçları Arasındaki İlişki	Logistic Regression, Veri Zarflama Analizi (Dea)	Algoritma %82.1 oranında doğruluk oranıyla tahminleme yapmıştır.
24	Mahmood vd	2021		Yükselen Yıldız Tahmini	Classification And Regression Trees (Cart), Support Vector Machine, Maximum Entropy Markov Model, Bayesian Network, Naive Bayes	F ölçümlerine göre CART ve SVM algoritmaları %89, Maximum Entropy Markov Model ve Naive Bayes algoritmaları %87 olarak tespit edilmiştir.
25	Wang ve Fan	2021		All-Star Tahmini, Playoff Tahmini ve Sıcak Seri Yanılgısı (Hot Streak Fallacy) Tahmini	K-Nearest Neighbor Classification, Logistic Regression, Decision Tree, Random Forest, Linear Regression ve Simple Regression	All-Star tahminleme K-Nearest Neighbor Classification %95.83, Logistic Regression %97.9 ve Decision Tree %95-%94 aralığında doğruluk oranı elde etmişlerdir. Playoff Tahmini için Random Forest %85 doğruluk oranıyla başarılı tahminleme yapmıştır. Sıcak seri için linear regression ve simple regression %40 doğruluk oranının altında kalmıştır.
26	Çene	2022	2016-2017 ve 2020-2021 arası	Maç Sonuçları	K-Nearest Neighbours, Naive Bayes, Lojistik Regresyon, SVM, Karar Ağaçları, Rastgele Orman ve Yapay Sinir Ağları	En iyi tahminleme yapanları Lojistik Regresyon %84.4, SVM %84.2 ve Yapay Sinir Ağları %84.5 doğruluk oranlarıyla tespit etmiştir. Savunma ribaundu, gerçek şut yüzdesi, top çalma, top kaybı, hücum ribaundu ve denenen serbest atış değişkenlerini sonuç tahmin modeli üzerinde en çok etki eden değişkenler olarak bulmuştur.
27	Nguyen vd.	2022		Oyuncularının All-Star Maçına Seçilip Seçilmeme Tahmini	Linear Regression, Gradient Boosting Machine, Support Vector Machine, Random Forest, Neural Network, Logistic Regression, Stochastic Gradient	En etkili değişken PTS (Total Points) olarak bulunmuştur. Dengeli alt örnekleme yapısıyla oluşturulan Random Forest modelinin en uygun algoritma olduğu tespit edilmiştir.

					Descent ve Naive Bayes	
28	Patrod vd.	2023		Maç Sonuçları	Logistic Regression, Naive Bayes Decision Trees, Multilayer Perceptron, K-Nearest Neighbours, Random Forest, Logitboost	K-nearest neighbours %57.9 doğruluk oranıyla en iyi çalışan algoritma olmuştur.
29	Papadaki ve Tsagris	2020	2017-2019 yılları arası	Maaş Tahmini	Losso Regresyon, Random Forest	2017-2018 ve 2018-2019 tahminleri için AUC değerleri sırasıyla 0.811 ve 0.841 olarak tespit edilmiştir ve cross-validation sürecinin sonuçlarını doğruladığını belirtmişlerdir.
30	William vd.	2020	2016-2017	Oyuncuların Free Agent (Serbest Oyuncu) Dönemindeki Maaş Tahmini	K-En Yakın Komşular	Ortalama hatası, 1.21'lik bir küme boyutu tespit edilmiştir. Düşük ve orta düzey oyunculardа hacim istatistiklerinin maaşlar için en belirleyici olduğu, yüksek VORP'un maksimum maaş alan tüm oyuncular arasında yaygın değişken olduğu ortaya çıkmıştır.
31	Özbalta vd.	2022	2019-2020	Maaş Tahmini	Regresyon ve Karar Ağacı	
32	Jain vd.	2022	2020-2021	Maaş Tahmini	Multiple Linear Regression, Decision Tree, Random Forest ve Xgboost	R^2oranları Multiple Linear Regression için %57.04, Decision Tree için %64.4, Random Forest için %75.85 ve Xgboost algoritması için % 88.18 olarak bulunmuştur.

DÖRDÜNCÜ BÖLÜM
MATERYAL VE YÖNTEM

4.1. Araştırmanın Materyal ve Yöntemi

Bu tez çalışmasında kullanılan makine öğrenmesi çalışmasının akış şeması Şekil 4.1'de gösterilmiştir.



Şekil 4. 1 Metodoloji Akış Şeması

4.2. Veri Toplama Süreci

Maaş tahminleme çalışması için gerekli olan oyuncuların istatistikleri (sayı ortalamaları, asist ortalamaları, rebound ortalamaları vb.), takımdaki rolü, yaş, draft sıralaması, sözleşme süresi, önceki maaşları, performans metrikleri gibi faktörler bu çalışma için önemlidir. Oyuncuların maaşlarına ilişkin güncel ve kapsamlı verilere erişim sağlanabilecek güvenilir kaynaklar, resmi NBA veri tabanı, spor istatistik web siteleri, NBA takımlarının web siteleri veya spor medyası gibi yerlerdir. Seçilen veri kaynağından oyuncuların maaşlarına ve ilgili özelliklere ilişkin verileri toplamak için manuel olarak veri girişi yapılabilecek yöntemler tercih edilmiştir.

NBA'in resmî sitesi olan Nba.com'dan 2021-2022 sezonunda yer alan tüm takımlar listelenmiştir (NBA, 2023). Takımların belirtilen sezondaki takım oyuncularının hepsi listelenmiştir. İlk listede çıkmayan oyuncuların 2021-2022 sezonundaki oynama sürelerine bakılarak sahada oynama süresi en yüksek olan oyuncular seçilmiştir.

NBA sitesinde oyuncular takımlara göre sıralanarak, sıradaki min (per game) filtrelemesiyle oyun başına oyunda oynama süresindeki büyükten küçüğe sıralanmıştır. İlk 8 sıradaki oyuncu bu tez kapsamında oluşturulan veri setine eklenmiştir. Bazı takımlarda bu sıralama yapıldığında 8 oyuncu olmadığı görülmüştür. Bu oyuncular NBA'in resmî sitesinde ilk 180 oyuncu arasında yer almadığı için 8 oyuncuyu tamamlayamadığımız takımların oyuncu kadrosundan en çok oynayan oyuncular hesaplanarak veri setine eklenmiştir.

Oyuncular belirlendikten sonra her oyuncu için basketball-reference.com sitesinden son sezon olan 2021-2022 sezonu istatistikleri, kariyer istatistikleri, Playoff istatistikleri, oyuncuların toplam ve son sezondaki Win Share puanları toplanmıştır (Basketball Reference, 2023). Win Share takımın başarısı için krediyi takımdaki oyunculara paylaştırmaya çalışan bir oyuncu istatistiğidir (Yang vd., 2021: 1405).

Tüm oyuncu performans istatistikleri maç başına ve her 36dk (per 36min) istatistikleri olarak iki ayrı veri seti oluşturulmuştur. NBA yıldızları maç başına en az 36 dakika oynadıkları için bu istatistik önemlidir.

WS (Win Share) değişkeni toplam ve son sezon verileri olmak üzere NBA'in resmî sitesinde yer alan her bir oyuncunun istatistiklerinde bulunan "Advanced" tablosundan çekilmiştir. Oyuncuların deneyimleri için olan Experience değişkeni de NBA'in sitesinden alınmıştır (NBA, 2023).

Veri toplarken Tablo 4.1'de yer alan oyuncuların Playoff'lara hiç katılmadığı görülmüştür. Bazı oyuncuların ise sezonda takım değiştirdikleri için son bulundukları takımların verileri alınmıştır.

Tablo 4.1 Playofflara Katılmayan Oyuncular ve Takımları

Oyuncu	**Takım**
Spencer Dinwiddie	DAL
Derrick White	BOS
Jevon Carter	MIL
James Harden	PHI
Nickeil Alexander-Walker	UTA
Seth Curry	BKN- BRK
Andre Drummond	BKN- BRK
Jevon Carter	MIL
Caris LeVert	CLE
Montrezl Harrell	CHA-CHO
Robert Covington	LUC
CJ McCollum	NOP
Ish Smith	WAS

Devin Vassell.	WAS
Domantas Sabonis	SAC
Justin Holiday	SAC
Trey Lyles	SAC
Tyrese Haliburton	IND
Buddy Hield	IND
Darius Bazley	OKC
Dennis Schroder	HOU
Donte DiVincenzo	SAC
Josh Hart	POR
Justise Winslow	POR
Jalen Smith	IND
Marvin Bagley III	DET

Oyuncuların maaş değişkeni olan Salary'ler 21.03.2023 tarihinde hoopshype web sitesinden alınmıştır (Hoopshype, 2023).

Takım istatistikleri için son sezonun verileri kullanılmıştır. Takımların Playoff istatistikleri basketball-reference web sitesinden alınmıştır (Basketball Reference, 2023). Takımlara ait Salary-Cap değerlerine Spotrac web sitesi üzerinden Total Cap ve Cap Space değerleri alınarak veri setine eklenmiştir (Spotrac, 2023). Takımlara ait NBA'e ait resmî web sitesinden yayınlanan sosyal medya hesapları yayınlanmaktadır. NBA'in resmi web sitesinde sadece Instagram ve Twitter sosyal medya platformları yer aldığı için sosyal medya olarak sadece bu iki platform kullanılmıştır. Instagram ve Twitter'dan takımların takipçi sayısı da veri setine eklenmiştir.

Ek 2'de veri setinde kullanılan değişkenlere ve karşılıklarına yer verilmiştir.

4.3. Veri Önişleme

Veri temizleme, bütünleştirme ve dönüştürme aşamaları bu aşamada yapılmıştır. Veri önişleme aşamaları, verilerin kalitesini artırarak, modelin güvenilir, doğru ve genellemesi iyi olan sonuçlar üretmesine yardımcı olmaktadır. Veri temizleme adımında toplanan NBA veri seti içerisinde eksik veya hatalı değerler, aykırı değerler ve tutarsızlıklar kontrol edilmiştir. Eksik verileri doldurma, aykırı değerleri ele alma veya tutarsızlık düzeltme gibi işlemlerle veri seti temizlenmiştir. Playoff'lara hiç katılmayan oyuncuların Playoff değişkenleri bulunmamaktaydı ve Playoff'la ilgili değişkenlerin olduğu sütunlarda veri seti boş kaldığı için ilgili alanlara sıfır değeri verilmiştir.

Veri Dönüştürme verileri makine öğrenmesi için uygun olan formata dönüştürme işlemidir (Oğuzlar, 2003: 73). Dönüştürme adımında NBA veri setinde farklı ölçeklerdeki

değişkenler bulunduğu tespit edilmiştir. Oyuncu boyu ve sayı ortalaması gibi değişkenler farklı aralıklarda bulunmakta olduğu görülmüştür. Bazı değerler tam sayı tipinde ve bazı değerler ise ondalıklı değerlerde olduğu gözlemlenmiştir. Bu durum bazı makine öğrenmesi modellerinin etkili bir şekilde çalışmasını engelleyebilmektedir. Veri ölçeklendirme yöntemleri ile (örneğin, min-max ölçeklendirme veya standartlaştırma) kullanarak değişkenler aynı ölçeğe indirilmiştir.

Öznitelik Seçimi NBA tahminlemesi yaparken, tüm özniteliklerin tahmin performansını artırmayabileceği durumlar olduğunda kullanılmıştır. Veri setindeki öznitelikleri analiz edip, modele en uygun olanları seçilmiştir. Öznitelik seçimi gereksiz veya düşük bilgi taşıyan öznitelikleri çıkarma ya da daha az boyutlu bir öznitelik değişkenlerine dönüştürme gibi işlemleri içerir. Veri seti oluşturulduğunda Tablo 4.2 'de yer alan toplam 98 değişken bulunmaktadır. Öznitelik seçiminde ID, Player, Team, Team-Abb, Salary_21-22, Salary_21-22_Current, Salary_22-23, WS_21-22 değişkenleri çıkartılmıştır. WS_Career değişkeni çıktı değişkeni olarak belirlenmiştir.

Veriyi Bölme aşamasında veri seti eğitim %20 ve test verisi %80 olmak üzere ikiye bölünmüştür. Eğitim verisi makine öğrenme modellerini eğitmek için kullanılmıştır. Test verisi ise modelin genel performansını değerlendirmek için kullanılmıştır.

Sosyal medya olan Instagram ve Twitter uygulamaları profillerin takipçi sayılarını kısaltma ve virgüllerle vermektedir. Sosyal medya hesaplarındaki oyuncuların takipçi sayısında yer alan "Billion" kelimesinin kısaltması olan "B" değerleri 1000 ile, "Million" kelimesinin kısaltması olan "M" değerleri 1000.000 ile çarpılmış ve takipçi sayılarındaki virgüller kaldırılmıştır. Takıma ait Salary Cap değişkenleri ve oyuncuların Salary değişkenlerinde yer alan "$" ve "," sembolleri kaldırılmıştır. Oyunculara ait deneyim değişkeni olan "Experience" de çaylak anlamına gelen Rookies'in kısaltması olarak R harfleri olduğu görülmüştür. Çaylakların deneyimi olmadığı için "R" harfi yerine 0 değeri verilmiştir. Oyuncu performanslarına ait FG%, 3P%, 2P%, eFG%, FT% değişkenlerinin değerlerinde yer alan "."ile başlayan sayısal değerler "0." olacak şekilde dönüştürülmüştür.

4.4. Makine Öğrenmesi İçin Kullanılan Bilgisayar ve Program

İşlemcisi Intel(R) Core(TM) i5-2400 CPU @ 3.10GHz 3.10 GHz 6,00 GB Ram kapasiteli bir bilgisayar kullanılmıştır. Bu tez çalışmasında makine öğrenimi uygulaması için görsel bir programlama arayüzüne sahip bir veri analizi ve makine öğrenimi platformu olan KNIME (Konstanz Information Miner) programı tercih edilmiştir (NBA Media Ventures, 2023b).

KNIME, makine öğrenimi için tercih edilen bir program olmasının birkaç nedeni vardır. KNIME görsel bir programlama arayüzü sağlar. Böylelikle kullanıcıların veri işleme, analiz ve makine öğrenimi işlemlerini sürükle-bırak yöntemiyle kolayca oluşturmasını sağlar. Kullanıcılar düğümleri birleştirerek veri akışını oluşturabilir, parametreleri ayarlayabilir ve sonuçları anında görebilir. Görsel arayüz, kullanıcıların karmaşık programlama becerilerine ihtiyaç duymadan hızlı bir şekilde işlem yapmalarına olanak tanır (NBA Media Ventures, 2023b).

Ayrıca KNIME programı çeşitli veri kaynaklarıyla entegrasyon sağlar. Excel, CSV, veritabanları, Hadoop, Spark ve daha fazlasını destekler. Farklı veri kaynaklarından veri alınmasına, birleştirilmesine ve analiz edilmesine olanak tanır.

KNIME bunların dışında makine öğrenimi ve veri analizi için çeşitli araçlar ve düğümler sunar. Sınıflandırma, regresyon, kümeleme, derin öğrenme ve daha fazlası gibi birçok makine öğrenimi algoritmasını içermektedir. Kullanıcılar bu düğümleri kullanarak modelleri oluşturabilir, eğitebilir ve değerlendirebilir. Ayrıca veri ön işleme, görselleştirme ve model değerlendirme gibi diğer veri analizi işlemleri de yapılabilmektedir.

Büyük veri işleme yeteneklerine sahip KNIME, Hadoop ve Spark gibi büyük veri platformlarını destekler. Büyük veri kümelerini işleyebilme, paralel hesaplama yapabilme ve veri analizi süreçlerini hızlandırabilme imkânı da sağlamaktadır.

Bunların dışında KNIME açık kaynaklı bir platformdur ve birçok kullanıcı tarafından geliştirilmektedir. Kullanıcıların kendi düğümlerini ve eklentilerini oluşturabilmesine ve KNIME topluluğunun gelişmiş işlevsellik ve özellikler sağlamasına olanak tanır. KNIME, kullanıcıların ihtiyaçlarına göre özelleştirilebilir ve genişletilebilir bir programdır. KNIME kullanıcı dostu arayüzü, geniş veri entegrasyonu, makine öğrenimi araçları ve genişletilebilirlik sağlayan bir program olarak makine öğrenimi projeleri için popüler bir seçenek haline gelmiştir.

4.5. Kullanılan CRISP-DM Yöntemi

Bu tez çalışmasının ikinci bölümün üçüncü alt başlığında anlatılan veri madenciliği metodolojisi olan CRISP-DM bu tez çalışmasında kullanılmıştır. Bu tez çalışmasında maaş tahmini yapmak için gerekli olan oyuncu ve takım bilgileri işi anlama aşamasında belirlenmiştir. Veriyi anlama aşamasında makine öğrenmesi tekniklerini uygulayabilmek için veri setinin oluşturulması gerekmektedir. İhtiyaç duyulan veriler basketball-reference ve nba.com/stats siteleri üzerinden ikincil kaynaklardan elde edilmiştir (Basketball Reference, 2023; NBA, 2023). 1950'ye kadar uzanan her NBA oyuncusu için yaklaşık 50 istatistik bulunmaktadır. Bu verilere maaş, oyuncu deneyimi, sosyal medya gibi maaşı etkileyeceği

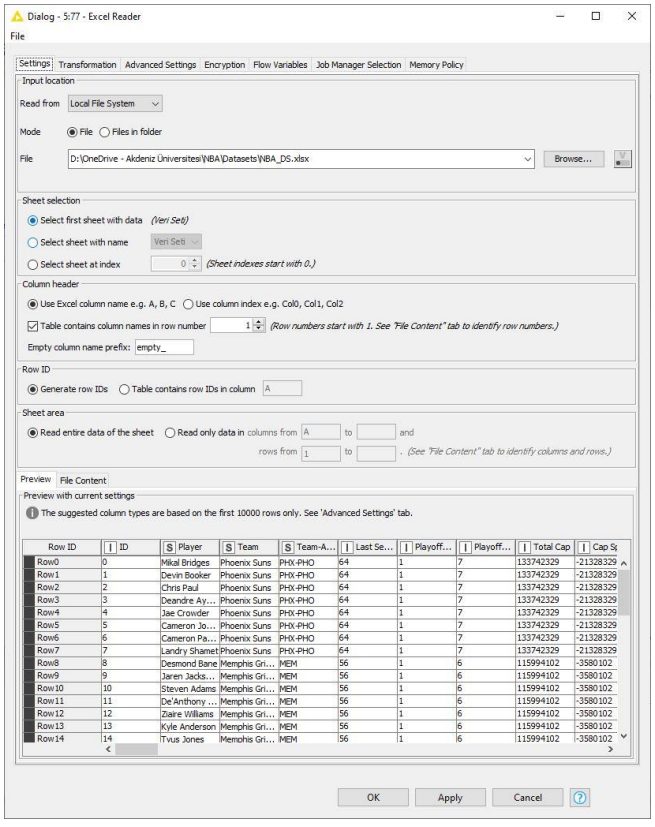
düşünülen değişkenler eklenmiştir. Veri seti elde edildikten sonra veriyi hazırlama aşaması için ön işleme adımları uygulanmıştır. Veri ön işleme adımlarında veri temizleme, veri standardizasyonu, veri indirgeme, değişken dönüşümü gibi işlemler yapılarak veri seti düzenlenmiştir. Veri seti ön işleme adımlarından sonra veri seti öğrenme ve test olarak ikiye ayrılmıştır. Modelleme aşamasında veri seti üzerinde yapılan düzenlemelerin ardından hangi makine öğrenimi algoritmalarının eğitim için kullanılacağı belirlemek için eğitim modelleri oluşturulmuştur. CRISP-DM metodolojisinde modellemeden sonra değerlendirme ve dağıtım aşamaları yer almaktadır. Bu aşamalara bulgular bölümünde yer verilecektir.

4.6. Makine Öğrenmesi Uygulaması

Bu aşamada KNIME Programında kullanılan düğümlerle model ayarlanmış ve hiper parametreleri seçilmiştir.

4.6.1. Excel Reader Düğümü

KNIME'da Excel Reader düğümü, Excel dosyalarını okumanıza ve içerdikleri verileri KNIME veri akışına aktarmanıza olanak tanır. Bu düğüm Excel dosyalarını farklı biçimlerde okuyabilir ve içerdikleri verileri tablo formatında sunar. Oluşturulan veri seti Excel formatında olduğu için Excel Reader düğümü kullanılarak okutulmuştur. Veri seti Şekil 4.2'de yer almaktadır.



Şekil 4.2 Excel Reader Ayarlar

Bunun için düğümde Configure-Settings-File'dan dosya yüklenmiştir. Excel dosyasındaki ilk satır değişken isimleri olduğu için default olarak gelen "Use Excel Column Name" seçili bırakıldı. Şekil 4.2'de Excel Reader ayarları yer almaktadır.

Düğümdeki File Table seçeneğinden verinin tamamı görüldü. Şekil 4.3'te görüleceği gibi veri seti 240 satırdan (gözlem) ve 98 sütundan (değişken) oluşmaktadır.

File Table - 5:77 - Excel Reader

File Edit Hilite Navigation View

Table "default" - Rows: 240 | Spec - Columns: 98 | Properties | Flow Variables

Row ID	ID	Player	Team	Team-A...	Last Se...	Playoff...	Playoff..
Row0	0	Mikal Bridges	Phoenix Suns	PHX-PHO	64	1	7
Row1	1	Devin Booker	Phoenix Suns	PHX-PHO	64	1	7
Row2	2	Chris Paul	Phoenix Suns	PHX-PHO	64	1	7
Row3	3	Deandre Ay...	Phoenix Suns	PHX-PHO	64	1	7

Şekil 4.3 Veri Seti

Excel Reader düğümünü kullanarak Şekil 4.4'te görülen tablodan veri keşfi yapılmıştır. En düşük ve yüksek değerler, tekrarlayan değerler incelenmiştir. Takımlarda sadece 30 değerin olduğu ve Pos değişkeninde 4 adet değer olduğu görülmüştür. Veri setindeki tüm değişkenlerden 4 tanesi String yani karakter veri tipinde, kalanı Double yani ondalıklı sayı olarak tespit edilmiştir.

File Table - 5:77 - Excel Reader

File

Table "default" - Rows: 240 | Spec - Columns: 98 | Properties | Flow Variables

Columns: 98	Column Type	Column Index	Color Handler	Size Handler	Shape Han...	Filter Handler	Lower Bound	Upper Bound	Value 0	Value 1	Value 2	Value 3	Value 4	Value 5	Value 6
ID	Number (integer)	0					0	239	?	?	?	?	?	?	?
Player	String	1					?	?	?	?	?	?	?	?	?
Team	String	2					?	?	Phoenix Suns	Memphis Gri...	Golden Stat...	Miami Heat	Dallas Mave...	Boston Celtics	Milwaukee B..
Team-Abb,	String	3					?	?	PHX-PHO	MEM	GSW	MIA	DAL	BOS	MIL
Last Season Win	Number (integer)	4					20	64	?	?	?	?	?	?	?
Playoffs katılma	Number (integer)	5					0	1	?	?	?	?	?	?	?
Playoffs Win	Number (integer)	6					0	16	?	?	?	?	?	?	?
Total Cap	Number (integer)	7					90,513,233	184,024,769	?	?	?	?	?	?	?
Cap Space	Number (integer)	8					-71,610,769	21,900,767	?	?	?	?	?	?	?
Total Cap (#1)	Number (integer)	9					101,549,048	200,320,856	?	?	?	?	?	?	?
Cap Space (#1)	Number (integer)	10					-76,665,856	22,105,952	?	?	?	?	?	?	?
InstagramTeam	Number (integer)	11					1,200,000	28,600,000	?	?	?	?	?	?	?
TwitterTeam	Number (integer)	12					1,100,000	11,400,000	?	?	?	?	?	?	?
Age	Number (integer)	13					19	37	?	?	?	?	?	?	?
Pos	String	14					?	?	SF	SG	PG	C	PF	?	?
G	Number (integer)	15					3	82	?	?	?	?	?	?	?
GS	Number (integer)	16					0	82	?	?	?	?	?	?	?
MP	Number (double)	17					9.9	40	?	?	?	?	?	?	?
FG	Number (double)	18					1.1	11.4	?	?	?	?	?	?	?
FGA	Number (double)	19					3.2	21.8	?	?	?	?	?	?	?
FG%	Number (double)	20					0.271	0.761	?	?	?	?	?	?	?
3P	Number (double)	21					0	4.5	?	?	?	?	?	?	?
3PA	Number (double)	22					0	11.7	?	?	?	?	?	?	?
3P%	Number (double)	23					0	10,000	?	?	?	?	?	?	?
2P	Number (double)	24					0.4	9.5	?	?	?	?	?	?	?
2PA	Number (double)	25					0.8	18.3	?	?	?	?	?	?	?
2P%	Number (double)	26					0.351	0.761	?	?	?	?	?	?	?
eFG%	Number (double)	27					0.281	0.761	?	?	?	?	?	?	?
FT	Number (double)	28					0.1	9.6	?	?	?	?	?	?	?
FTA	Number (double)	29					0.2	11.8	?	?	?	?	?	?	?
FT%	Number (double)	30					0.392	10,000	?	?	?	?	?	?	?
ORB	Number (double)	31					0.1	4.6	?	?	?	?	?	?	?
DRB	Number (double)	32					1.3	11	?	?	?	?	?	?	?
TRB	Number (double)	33					1.5	14.7	?	?	?	?	?	?	?
AST	Number (double)	34					0.3	10.8	?	?	?	?	?	?	?
STL	Number (double)	35					0.3	2	?	?	?	?	?	?	?
BLK	Number (double)	36					0	2.8	?	?	?	?	?	?	?

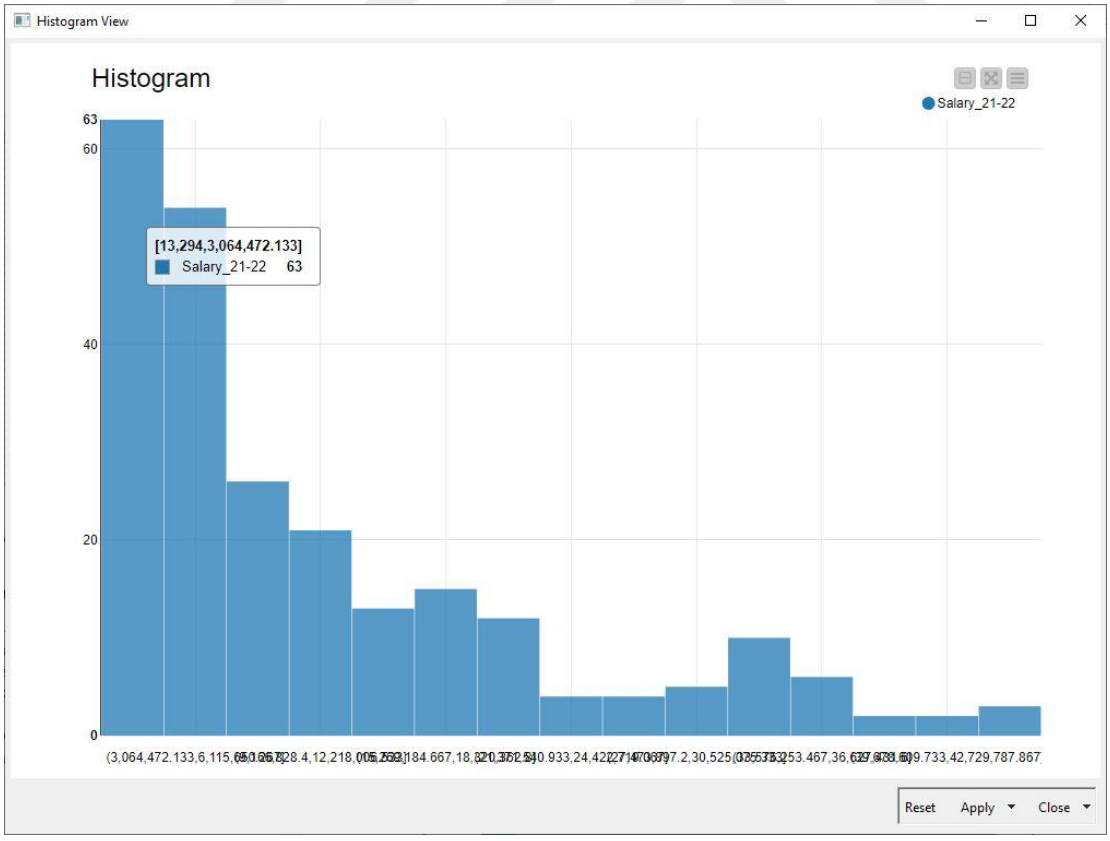
Şekil 4.4 Excel Reader Düğümü Değişkenler

Veri setimizde çıktı değişkeni olduğu için Denetimli Öğrenme kullanılmıştır. Kullanılan veri setindeki çıktı değişkeni bir sayı değişkeni olduğu için bir değişiklik yapılmamıştır.

4.6.2. Histogram Düğümü

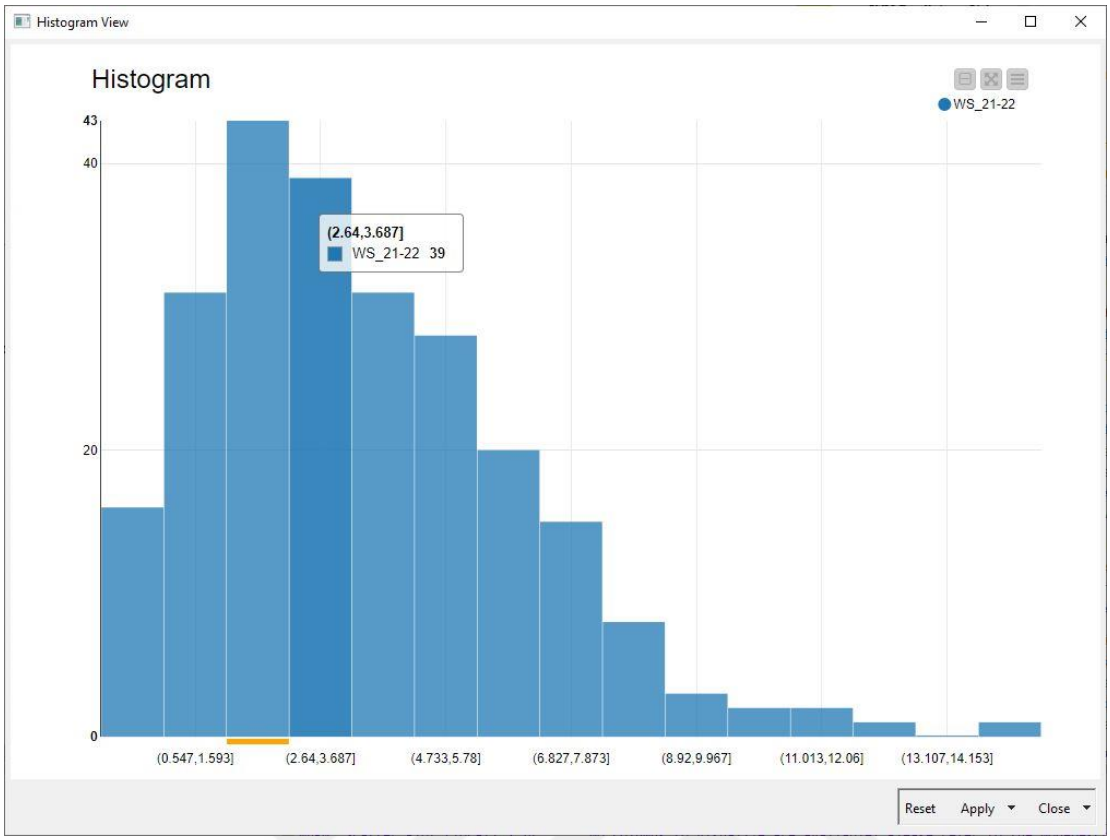
Histogram düğümü KNIME'da veri setlerinin histogramlarını oluşturmak için kullanılan bir düğümdür. Veri setinin dağılımını daha iyi anlamak ve verilerin yoğunlukları hakkında bilgi edinmek için kullanılmıştır. Histogramlar, veri setindeki değerlerin dağılımını görselleştirerek veri analizini kolaylaştırır. Bu sayede veri setindeki önemli desenleri ve eğilimleri tespit etmek ve verilerin daha iyi anlaşılmasına yardımcı olması için kullanılmıştır. Histogram tek değişkenli bir analizdir. Ayarlar-Binning kısmında kaç tane çubuk olacağını belirlemek için "Number of Bins" değeri 15 olarak belirlenmiştir. "Bin Naming" borders seçeneği daha çok bilgi almak için seçilmiştir. Border sayesinde her kategorinin minimum ve maximumları görülmüştür. Salary_21-22 değişkeni için histograma bakıldığında ilk çubuk 13,294 ve 3.064,472.133 Salary_21-22 değerleri arasındakileri anlatmaktadır. Veride 13,294 ve 3.064,472.133 Salary_21-22 değerleri arasında 63 gözlem vardır. Şekil 4.5'te Salary_21-22 tüm aralıklar için incelendiğinde hepsinde farklı gözlem sayısı için geçerli olduğu görülmektedir. İlk iki aralığın diğerlerine oranla daha çok gözlem içerdiği görülebilmektedir.

Bütün değişkenler için histogramları incelendiğinde sadece win share değişkenleri normal dağılım göstermiştir.



Şekil 4.5 Histogram Salary 21-22

Şekil 4.6'de WS_21-22 değişkeni için histograma bakıldığında tüm aralıklar için incelendiğinde normal dağılıma daha yakın olduğu görülmektedir.



Şekil 4.6 Histogram WS 21-22

4.6.3. Statistics Düğümü

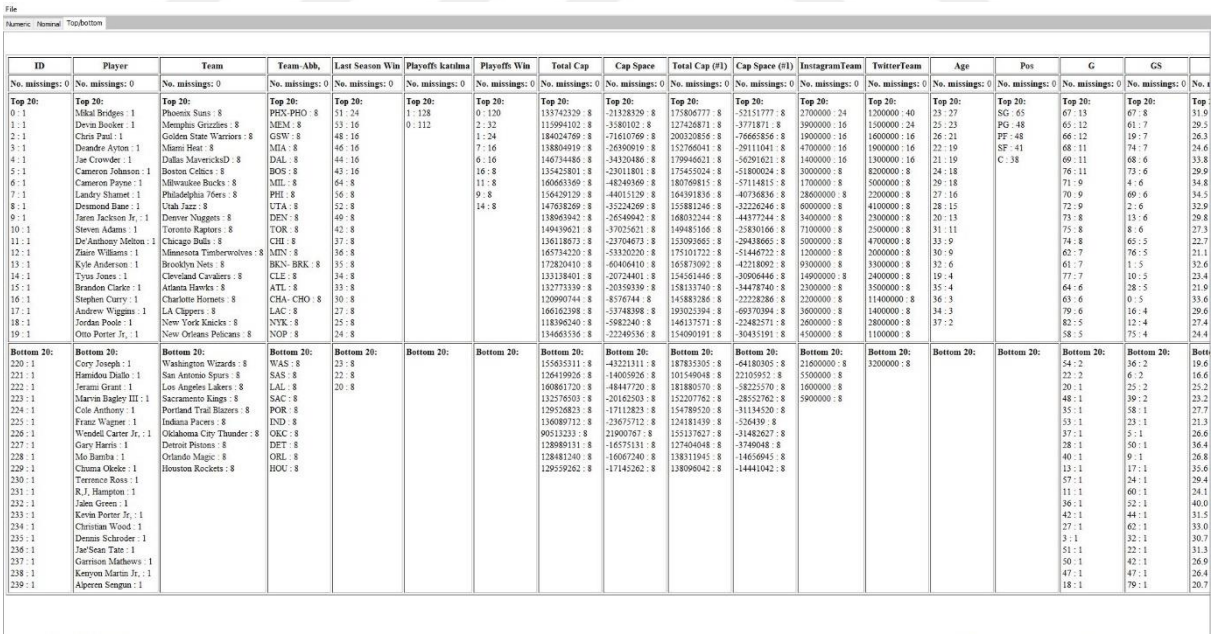
Statistics düğümü istatistiksel analizler yapmak için kullanılan bir düğümdür. Bu düğümde veri setinin temel tanımlayıcı istatistik bilgilerine bakılmıştır. Last Season Win değişkeninin minimum değeri 20, orta değeri 41, maximum değeri 64, standart sapması 11,3962 olarak görülmektedir. Skewness ve kurtosis çarpıklıkla ilgilidir. Statistic düğümünde her değişken için temel değerler birlikte görülmektedir. Hiçbir değişkende eksik veri olmadığı gözlemlenmiştir (Şekil 4.7).



Column	Min	Mean	Median	Max	Std. Dev.	Skewness	Kurtosis	No. Missing	No. +∞	No. -∞	Histogram
ID	0.0	119,5	?	239	69,4262	0.0	-1,2	0	0	0	
Last Season Win	20	41	?	64	11,3962	-0,2611	-0,8658	0	0	0	
Playoffs katılma	0.0	0,5333	?	1	0,4999	-0,1345	-1,9986	0	0	0	
Playoffs Win	0.0	2,9	?	16	4,4466	1,6056	1,5425	0	0	0	
Total Cap	90.513.233	1,40E8	?	1,84E8	18.939.084,2729	0,0705	0,4577	0	0	0	
Cap Space	-71.610.769	-27.829.718,6	?	21.900.767	18.939.084,2729	-0,0705	0,4577	0	0	0	
Total Cap (#1)	1,02E8	1,58E8	?	2,00E8	21.706.360,7162	-0,3087	0,0797	0	0	0	
Cap Space (#1)	-76.665.856	-34.264.197	?	22.105.952	21.706.360,7162	0,3087	0,0797	0	0	0	
InstagramTeam	1.200.000	5.396.666,6667	?	28.600.000	6.015.126,4463	2,6519	6,642	0	0	0	

Şekil 4.7 Statistics Düğümü

Şekil 4.8'de değişkenlerin taban ve tavan değerleri yer almaktadır. Pos, Playoffs katılma, Team gibi kategorik değişkenlerin her bir kategorisi için azalan gösterimde kaç gözlem olduğu bu tablodan görülmektedir. Örneğin Pos değişkeninde SG için 65, PG için 48, PF için 48, SF için 41 ve C kategorisi için 38 gözlem bulunmaktadır.



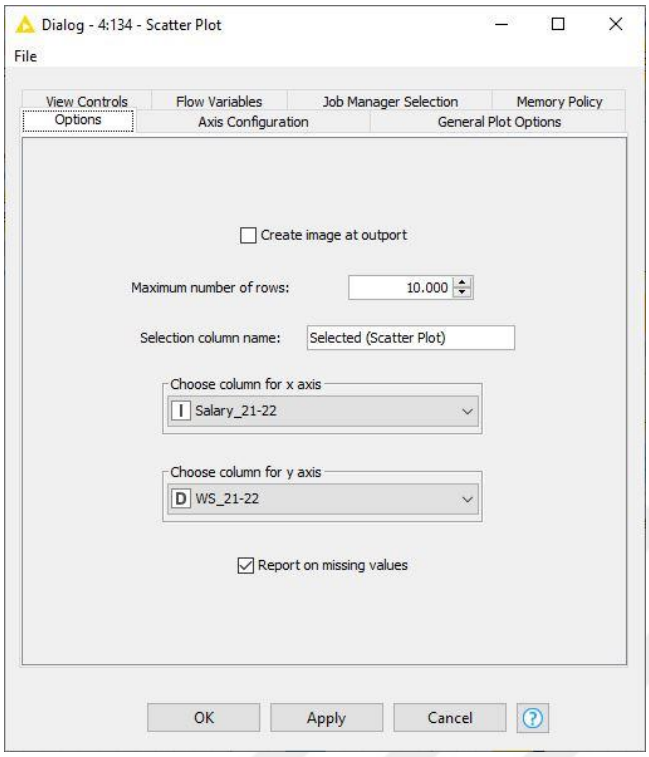
ID	Player	Team	Team-Abb,	Last Season Win	Playoffs katılma	Playoffs Win	Total Cap	Cap Space	Total Cap (#1)	Cap Space (#1)	InstagramTeam	TwitterTeam	Age	Pos	G	GS	
No. missings: 0	No. missings: 0	No. missings: 0	No. missings: 0	No. missings: 0	No. missings: 0	No. missings: 0	No. missings: 0	No. missings: 0	No. missings: 0	No. missings: 0	No. missings: 0	No. missings: 0	No. missings: 0	No. missings: 0	No. missings: 0	No. missings: 0	
Top 20:	Top 20:	Top 20:	Top 20:	Top 20:	Top 20:	Top 20:	Top 20:	Top 20:	Top 20:	Top 20:	Top 20:	Top 20:	Top 20:	Top 20:	Top 20:	Top 20:	
0 : 1	Mikal Bridges : 1	Phoenix Suns : 8	PHX-PHO : 8	51 : 24	1 : 128	0 : 120	153742329 : 8	-21328329 : 8	175806777 : 8	-52151777 : 8	2700000 : 24	1200000 : 40	23 : 27	SG : 65	67 : 13	67 : 8	31.9
1 : 1	Devin Booker : 1	Memphis Grizzlies : 8	MEM : 8	48 : 16	0 : 112	2 : 32	115994102 : 8	-3580102 : 8	127426871 : 8	-3771871 : 8	3900000 : 16	1500000 : 24	25 : 23	PG : 48	65 : 12	61 : 7	29.5
2 : 1	Chris Paul : 1	Golden State Warriors : 8	GSW : 8	48 : 16		1 : 24	184024769 : 8	-71610769 : 8	200320856 : 8	-76665856 : 8	1900000 : 16	1600000 : 16	26 : 21	PF : 48	66 : 12	19 : 7	26.3
3 : 1	Deandre Ayton : 1	Miami Heat : 8	MIA : 8	46 : 16		7 : 16	138804919 : 8	-26390919 : 8	152766041 : 8	-29111041 : 8	4700000 : 16	1900000 : 16	22 : 19	SF : 41	68 : 11	74 : 7	24.6
4 : 1	Jae Crowder : 1	Dallas MavericksD : 8	DAL : 8	44 : 16		6 : 16	146734486 : 8	-34320486 : 8	179946621 : 8	-36291621 : 8	1400000 : 16	1300000 : 16	21 : 19	C : 38	69 : 11	68 : 6	33.8
5 : 1	Cameron Johnson : 1	Boston Celtics : 8	BOS : 8	43 : 16		16 : 8	135425801 : 8	-23011801 : 8	175465024 : 8	-51800024 : 8	3000000 : 8	8200000 : 8	24 : 18		76 : 11	73 : 6	29.9
6 : 1	Cameron Payne : 1	Milwaukee Bucks : 8	MIL : 8	64 : 8		11 : 8	160663569 : 8	-48246369 : 8	180769815 : 8	-57114815 : 8	1700000 : 8	5000000 : 8	29 : 18		71 : 9	4 : 6	34.8
7 : 1	Landry Shamet : 1	Philadelphia 76ers : 8	PHI : 8	56 : 8		9 : 8	156439129 : 8	-44015129 : 8	164391836 : 8	-40736836 : 8	28600000 : 8	2200000 : 8	27 : 16		70 : 9	69 : 6	34.5
8 : 1	Desmond Bane : 1	Utah Jazz : 8	UTA : 8	52 : 8		14 : 8	147638269 : 8	-35224269 : 8	155881246 : 8	-32226246 : 8	6000000 : 8	4100000 : 8	28 : 15		72 : 9	2 : 6	32.9
9 : 1	Jaren Jackson Jr. : 1	Denver Nuggets : 8	DEN : 8	49 : 8			138963942 : 8	-26540942 : 8	168032244 : 8	-44377244 : 8	3400000 : 8	2300000 : 8	20 : 13		73 : 8	13 : 6	29.8
10 : 1	Steven Adams : 1	Toronto Raptors : 8	TOR : 8	42 : 8			149439621 : 8	-37025621 : 8	149485166 : 8	-25830166 : 8	7100000 : 8	2500000 : 8	31 : 11		75 : 8	8 : 6	27.3
11 : 1	De'Anthony Melton : 1	Chicago Bulls : 8	CHI : 8	37 : 8			136118673 : 8	-23704673 : 8	153093685 : 8	-29438665 : 8	5000000 : 8	4700000 : 8	33 : 9		74 : 8	65 : 5	22.7
12 : 1	Zaire Williams : 1	Minnesota Timberwolves : 8	MIN : 8	36 : 8			165734220 : 8	-53320220 : 8	175101722 : 8	-51446722 : 8	1200000 : 8	2000000 : 8	30 : 9		62 : 7	76 : 5	21.1
13 : 1	Kyle Anderson : 1	Brooklyn Nets : 8	BKN- BRK : 8	35 : 8			172820410 : 8	-60406410 : 8	165873092 : 8	-42218092 : 8	9300000 : 8	3300000 : 8	32 : 6		61 : 7	1 : 5	32.6
14 : 1	Tyus Jones : 1	Cleveland Cavaliers : 8	CLE : 8	34 : 8			133138401 : 8	-20724401 : 8	154561446 : 8	-30906446 : 8	14900000 : 8	2400000 : 8	19 : 4		77 : 7	10 : 5	23.4
15 : 1	Brandon Clarke : 1	Atlanta Hawks : 8	ATL : 8	33 : 8			132773339 : 8	-20359339 : 8	158133740 : 8	-34478740 : 8	2300000 : 8	3500000 : 8	35 : 4		64 : 6	28 : 5	21.9
16 : 1	Stephen Curry : 1	Charlotte Hornets : 8	CHA- CHO : 8	30 : 8			120990744 : 8	-8576744 : 8	145883286 : 8	-22228286 : 8	2200000 : 8	11400000 : 8	36 : 3		63 : 6	0 : 5	33.6
17 : 1	Andrew Wiggins : 1	LA Clippers : 8	LAC : 8	27 : 8			166162398 : 8	-33748398 : 8	193023394 : 8	-69370394 : 8	3600000 : 8	1400000 : 8	34 : 3		79 : 6	16 : 4	29.6
18 : 1	Jordan Poole : 1	New York Knicks : 8	NYK : 8	25 : 8			118396240 : 8	-5882240 : 8	140137571 : 8	-22482571 : 8	2600000 : 8	2800000 : 8	37 : 2		82 : 5	12 : 4	27.4
19 : 1	Otto Porter Jr. : 1	New Orleans Pelicans : 8	NOP : 8	24 : 8			134683536 : 8	-22249536 : 8	154090191 : 8	-30435191 : 8	4500000 : 8	1100000 : 8			58 : 5	75 : 4	24.4
Bottom 20:	Bottom 20:	Bottom 20:	Bottom 20:	Bottom 20:	Bottom 20:	Bottom 20:	Bottom 20:	Bottom 20:	Bottom 20:	Bottom 20:	Bottom 20:	Bottom 20:	Bottom 20:	Bottom 20:	Bottom 20:	Bottom 20:	Bott
220 : 1	Cory Joseph : 1	Washington Wizards : 8	WAS : 8	23 : 8			155635311 : 8	-43221311 : 8	187835305 : 8	-64180305 : 8	21600000 : 8	3200000 : 8			24 : 2	36 : 2	19.6
221 : 1	Hamidou Diallo : 1	San Antonio Spurs : 8	SAS : 8	22 : 8			126419926 : 8	-14005926 : 8	101549048 : 8	22105952 : 8	5500000 : 8				22 : 2	6 : 2	16.6
222 : 1	Jerami Grant : 1	Los Angeles Lakers : 8	LAL : 8	20 : 8			160861720 : 8	-48447720 : 8	181880570 : 8	-58225570 : 8	1600000 : 8				20 : 1	25 : 2	25.2
223 : 1	Marvin Bagley III : 1	Sacramento Kings : 8	SAC : 8				132576503 : 8	-20162503 : 8	152207762 : 8	-28552762 : 8	5900000 : 8				48 : 1	39 : 2	23.2
224 : 1	Cole Anthony : 1	Portland Trail Blazers : 8	POR : 8				129526823 : 8	-17112823 : 8	154789520 : 8	-31134520 : 8					35 : 1	58 : 1	27.7
225 : 1	Franz Wagner : 1	Indiana Pacers : 8	IND : 8				136089712 : 8	-23675712 : 8	124181439 : 8	-526439 : 8					53 : 1	23 : 1	21.3
226 : 1	Wendell Carter Jr. : 1	Oklahoma City Thunder : 8	OKC : 8				90513233 : 8	21900767 : 8	155137627 : 8	-31482627 : 8					5 : 1	5 : 1	26.6
227 : 1	Gary Harris : 1	Detroit Pistons : 8	DET : 8				128989131 : 8	-16575131 : 8	127404048 : 8	-3749048 : 8					28 : 1	50 : 1	36.4
228 : 1	Mo Bamba : 1	Orlando Magic : 8	ORL : 8				128481240 : 8	-16067240 : 8	138311945 : 8	-14656945 : 8					40 : 1	9 : 1	26.8
229 : 1	Chuma Okeke : 1	Houston Rockets : 8	HOU : 8				129559262 : 8	-17145262 : 8	138096042 : 8	-14441042 : 8					13 : 1	17 : 1	35.6
230 : 1	Terrence Ross : 1														57 : 1	24 : 1	29.4
231 : 1	R.J. Hampton : 1														11 : 1	60 : 1	24.1
232 : 1	Jalen Green : 1														36 : 1	52 : 1	40.0
233 : 1	Kevin Porter Jr. : 1														42 : 1	44 : 1	31.5
234 : 1	Christian Wood : 1														27 : 1	62 : 1	33.0
235 : 1	Dennis Schroder : 1														3 : 1	32 : 1	30.7
236 : 1	Jae'Sean Tate : 1														51 : 1	22 : 1	31.3
237 : 1	Garrison Mathews : 1														50 : 1	42 : 1	26.9
238 : 1	Kenyon Martin Jr. : 1														47 : 1	47 : 1	26.4
239 : 1	Alperen Sengun : 1														18 : 1	79 : 1	20.7

Şekil 4.8 Statistics Düğümü Top/Bottom

4.6.4. Scatter Plot (Dağılım/Serpilme Grafiği) Düğümü

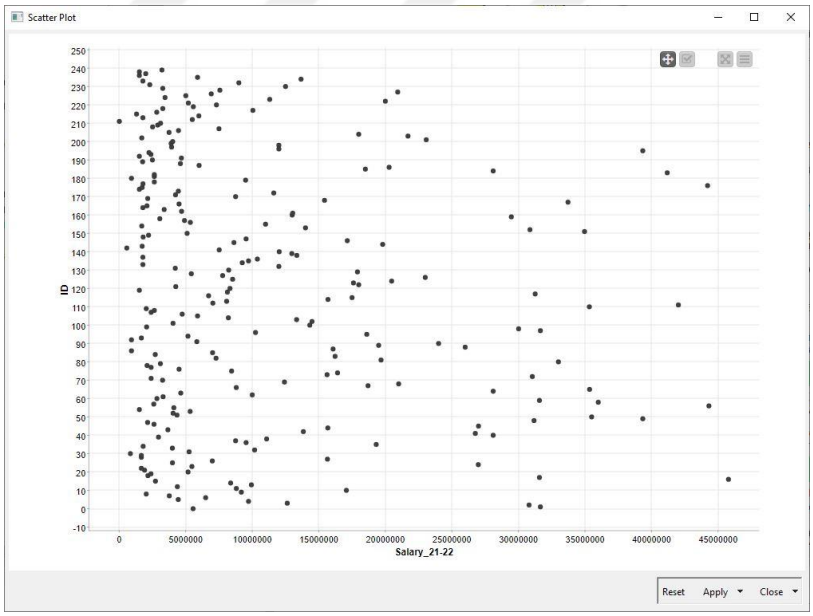
Scatter plot, veri noktalarının bir düzlemde dağılımını göstermek için kullanılan bir grafik türüdür. Scatter plot ile iki değişken arasındaki ilişkiyi veya dağılımı görselleştirmek için kullanılmıştır. Veri bu düğüm ile detaylı incelenmiştir. "Maximum number of rows" değeri tüm

gözlemler için uygulansın diye 10000'e çıkartılmıştır. Salary_21-22 değişkenini x eksenine, ID değişkenini y eksenine olacak şekilde ayarladığımızda Şekil 4.9'deki gibi ayarlanmıştır.



Şekil 4.9 Scatter Plot Düğümü

Her bir oyuncu için maaş dağılımı Şekil 4.10'de yer almaktadır.

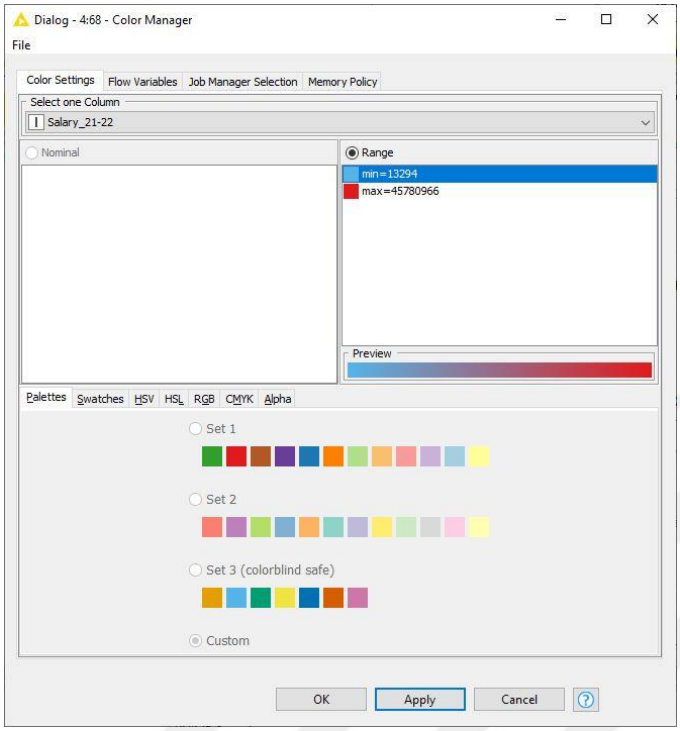


Şekil 4.10 Scatter Plot Maaş Oyuncu Grafiği

4.6.5. Color Manager Düğümü

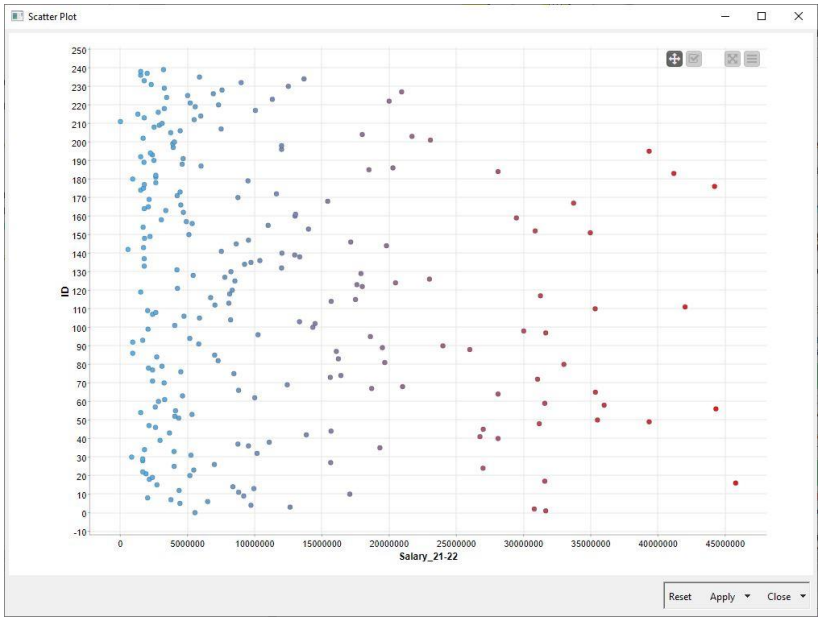
Color Manager düğümü, KNIME'da renklerin analiz ve görselleştirme işlemlerinde kullanılmasını kolaylaştırmak için kullanılan bir düğümdür. Scatter plottaki bilgiyi daha iyi

görebilmek ve daha rahat yorumlayabilmek için Color Manager düğümü kullanılmıştır. Veri seti önce Color Manager sonra Scatter Plot düğümüne bağlanmıştır. Color Manager ayarlarında Salary_21-22'nin en düşük değeri mavi en yüksek değeri kırmızı olacak şekilde renklendirilmiştir. Şekil 4.11'da Color Manager düğüm ayarları yer almaktadır.



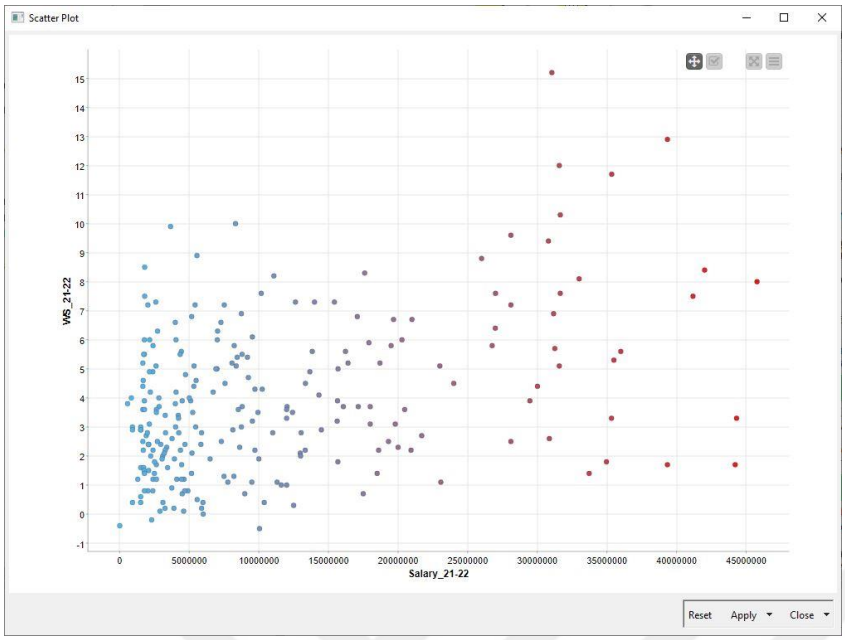
Şekil 4.11 Color Manager Düğümü

Her bir oyuncu için maaş dağılımı Şekil 4.12'de yer almaktadır. Maaşların 0 ve 10000000 arasında daha yoğun serpilme yaptığı görülmüştür. Bu yayılımlar tüm değişkenler için incelenmiştir. Bazı değişkenler için yayılım grafiklerine yer verilmiştir.



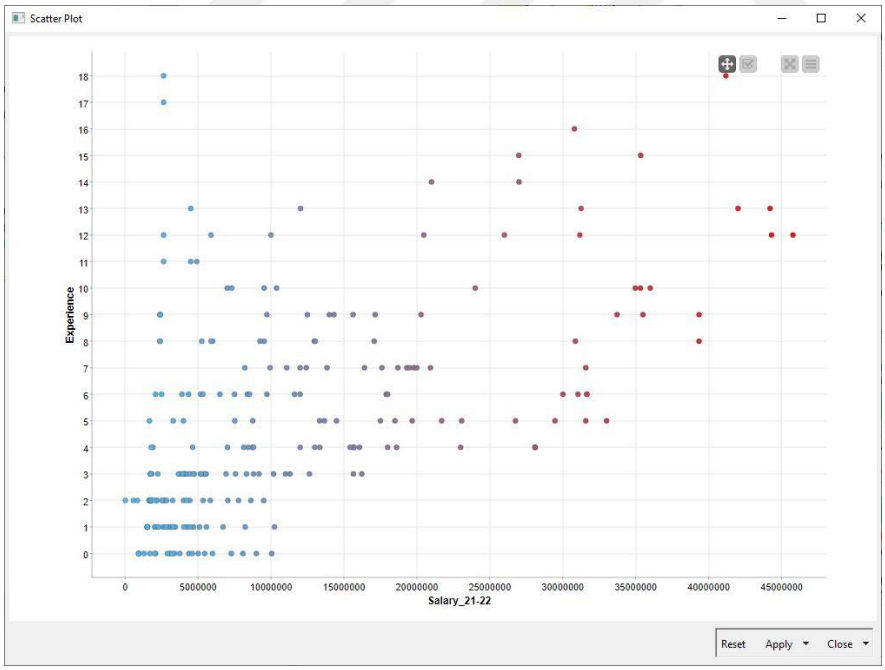
Şekil 4.12 Scatter Plot Maaş ve ID

Maaş ve aynı sezondaki WS ilişkisini gösteren grafik aşağıdaki Şekil 4.13'da yer almaktadır.



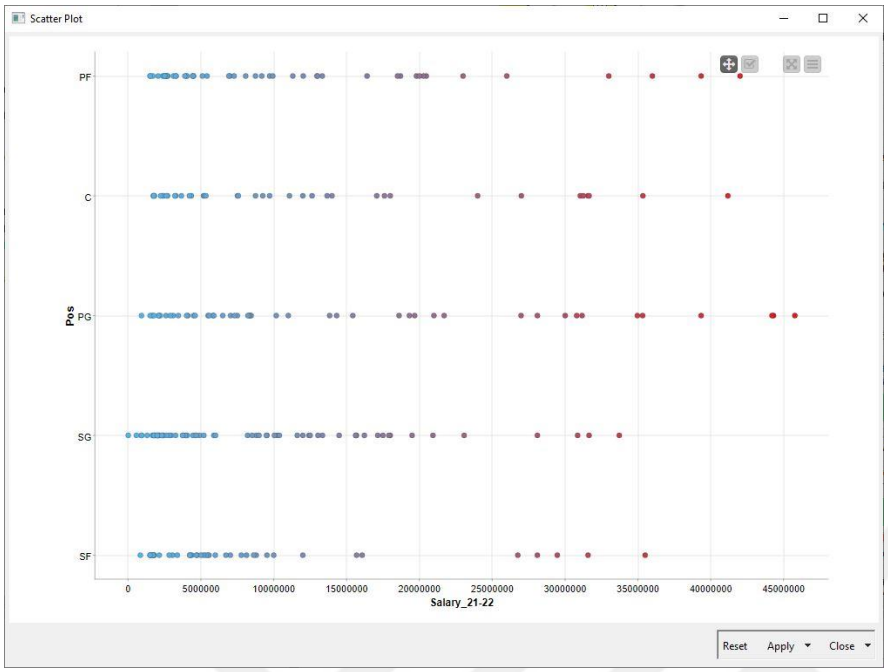
Şekil 4.13 Scatter Plot Maaş ve WS

Deneyim ve maaş arasındaki ilişkiyi gösteren grafik aşağıdaki Şekil 4.14'de yer almaktadır.



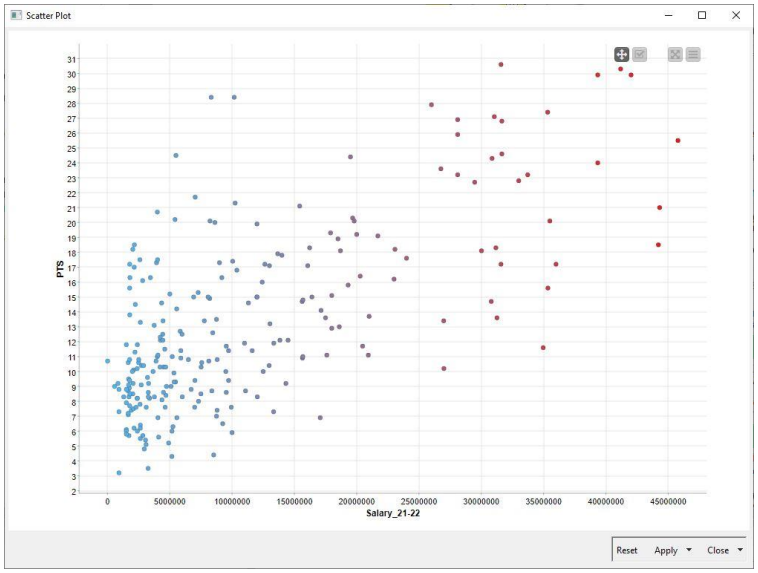
Şekil 4.14 Scatter Plot Maaş ve Deneyim

Maaşların oyuncu pozisyonlarıyla ilişkisi aşağıdaki Şekil 4.15'de gösterilmiştir. Grafik incelendiğinde maaşlara göre herhangi bir pozisyon kategorisi öne çıkmamıştır.



Şekil 4.15 Scatter Plot Maaş ve Pozisyon

Maaşların ile PTS (Önceki Sezon Maç Başına Attığı Ortalama Sayı) ile ilişkisi aşağıdaki Şekil 4.16'de gösterilmiştir. Bu grafikte maaşla PTS arasında doğrusal bir şekilde artıştan söz edilebilir.

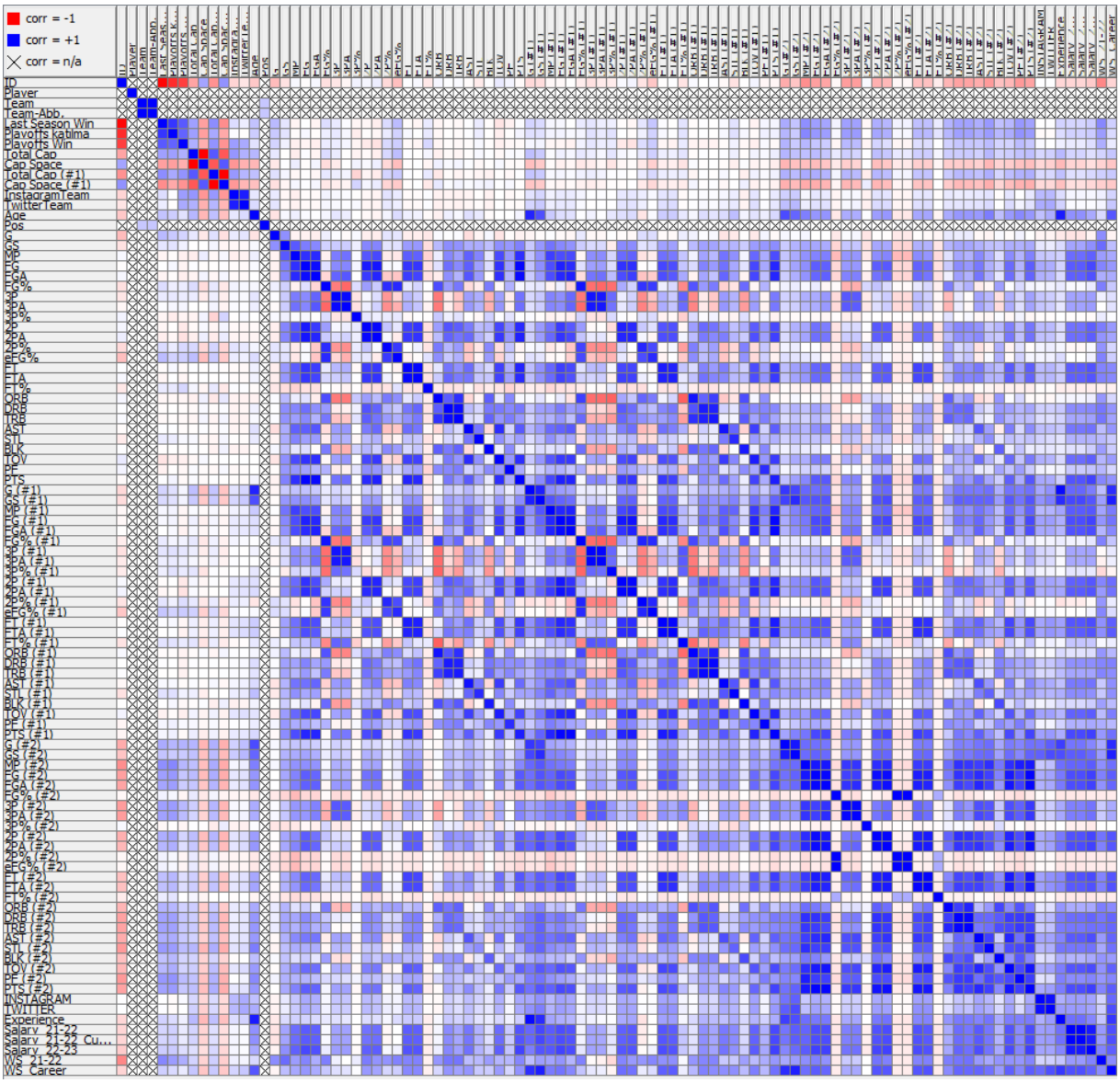


Şekil 4.16 Scatter Plot Maaş ve PTS

4.6.6. Linear Correlation Düğümü

Korelasyonlar değişkenler arasındaki ilişkiyi ve yönünü gösterir. Lineer korelasyon, iki değişken arasındaki doğrusal ilişkiyi ölçer. Örneğin, bir değişkenin artmasıyla diğer değişkenin de artması veya azalması durumunda, bu değişkenler arasında pozitif bir lineer korelasyon

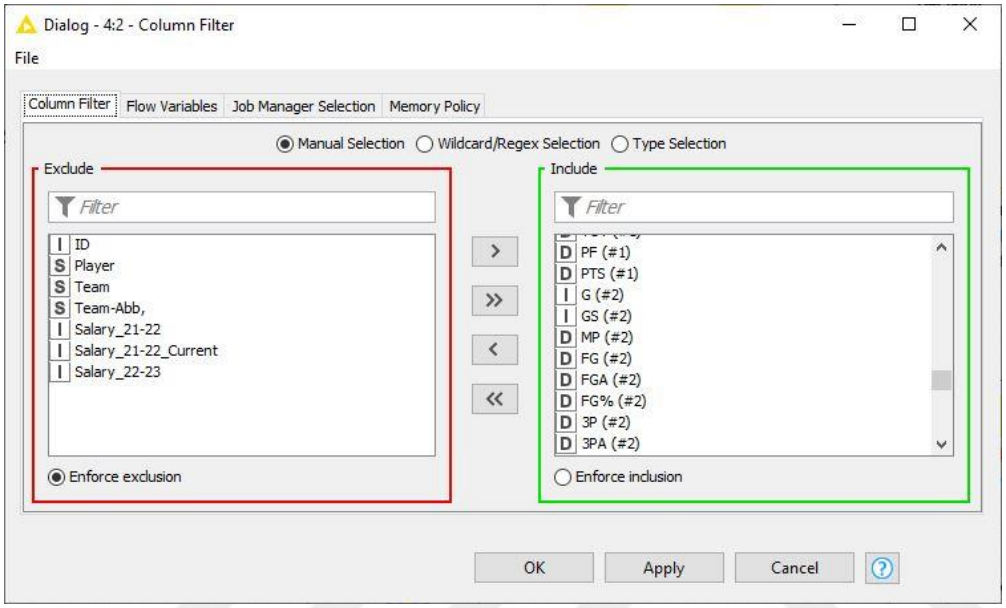
olduğu söylenebilmektedir. Yüksek korelasyon içeren değişkenleri birlikte kullanmaktan kaçınılmıştır. En yüksek Salary_21-22 ile FG, FGA, FT, FTA, PTS, MP(1), 2P(1), 2PA(1), FT(1), FTA(1), TOV(1), GS(2), MP(2), 2P(2), FT(2), FTA(2), AST(2), TOV(2), PTS(2), WS_Career değişkenleri arasında 0,60'dan yüksek korelasyon vardır. Salary_21-22 ile GS(1), FG(1), FGA(1), PTS(1), FG(2), FGA(2), 2PA(2) değişkenleri arasında 0,70-0,80 arasında korelasyon vardır. Salary_21-22 ile Cap Space ve Cap Space(1) arasında negatif yönlü bir korelasyon vardır. WS_21-22 değişkeni ile %60'dan yüksek korelasyonu olan bir değişken yoktur. Bu korelasyonlar çok yüksek olmadığı için değişkenlerin birlikte değerlendirilmesinde bir sakınca olmadığı düşünülmektedir. Şekil 4.17'da değişkenler arası Lineer Korelasyon tablosu yer almaktadır.



Şekil 4.17 Lineer Correlation

4.6.7. Column Filter

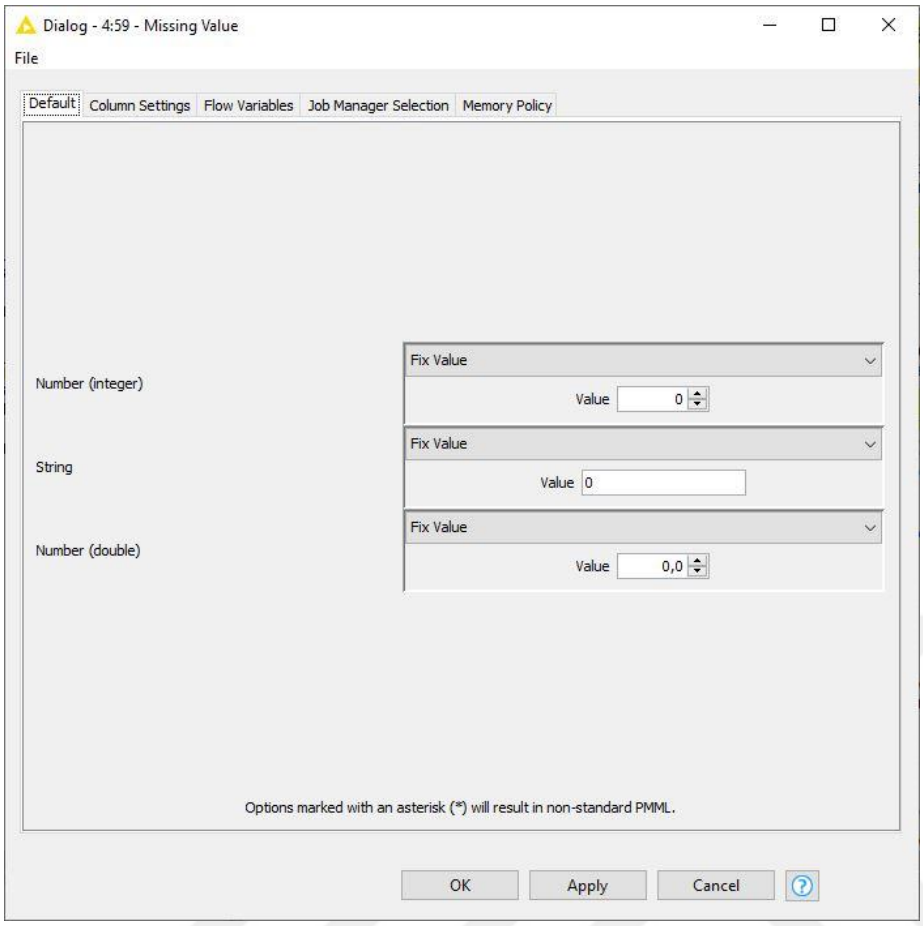
Bu düğüm veri setinde kullanılmak istenmeyen değişken sütunlarını filtrelemek için kullanılmaktadır. ID, Player, Team, Team-Abb, Salary_21-22, Salary_21-22_Current, Salary_22-23, WS_21-22 değişkenleri modele dahil edilmemek için filtrelenerek veri setinden çıkartılmıştır. WS_Career değişkeni çıktı değişkeni olacağı için modele diğer değişkenlerle birlikte dahil edilmiştir. Şekil 4.18'de Column Filter düğüm ayarları yer almaktadır.



Şekil 4.18 Column Filter Düğümü

4.6.8. Missing Value Düğümü

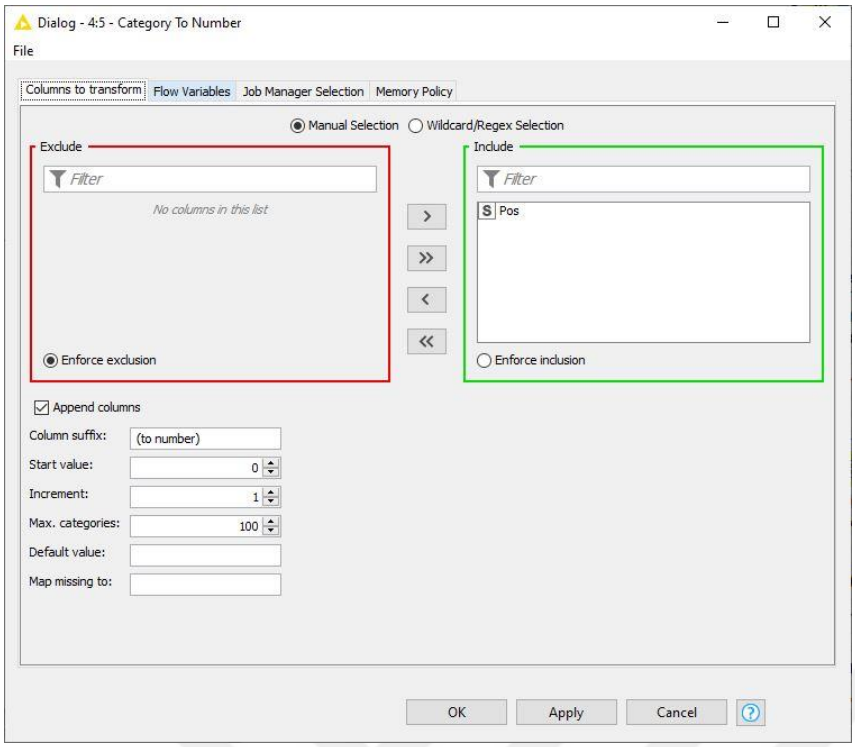
Missing Value düğümü, girdi tablosunun hücrelerinde bulunan eksik değerlere işlem yapılmasına yardımcı olur. Veri setinde bulunan eksik değerlerin yerine 0 sayısı verilecek şekilde düğüm ayarları yapılmıştır. Şekil 4.19'de düğüm ayarları gösterilmiştir.



Şekil 4.19 Missing Value Düğümü

4.6.9. Category to Number

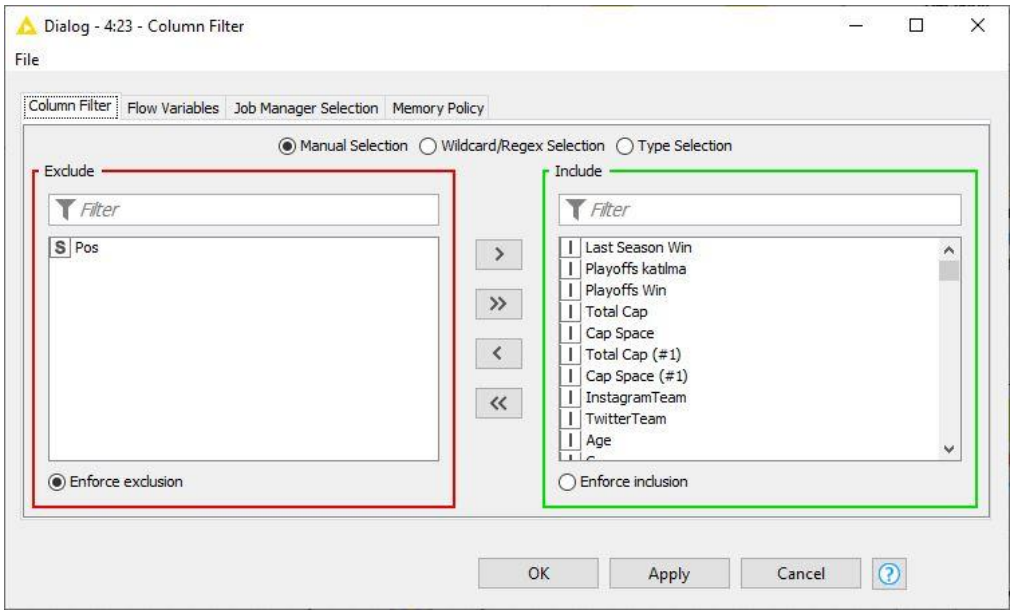
Category to Number düğümü kategorik verileri sayısal değerlere dönüştürmek için kullanılan bir işlem düğümüdür. Veri setindeki kategorik değişken olan Pos değişkeninin sayı karşılıklarını gösterebilmek amacıyla bu düğüm kullanılmıştır. Bu düğüm yeni bir Pos değişkeni oluşturarak, her bir kategoriye karşılık gelen sayısal olarak 0 ve 100 sayıları arasında karşılık değerini verecek şekilde yeni bir Pos değişkeni oluşturacak şekilde ayarlanmıştır. Şekil 4.20'da düğüm ayarları gösterilmiştir.



Şekil 4.20 Category to Number Düğümü

4.6.10. Column Filter

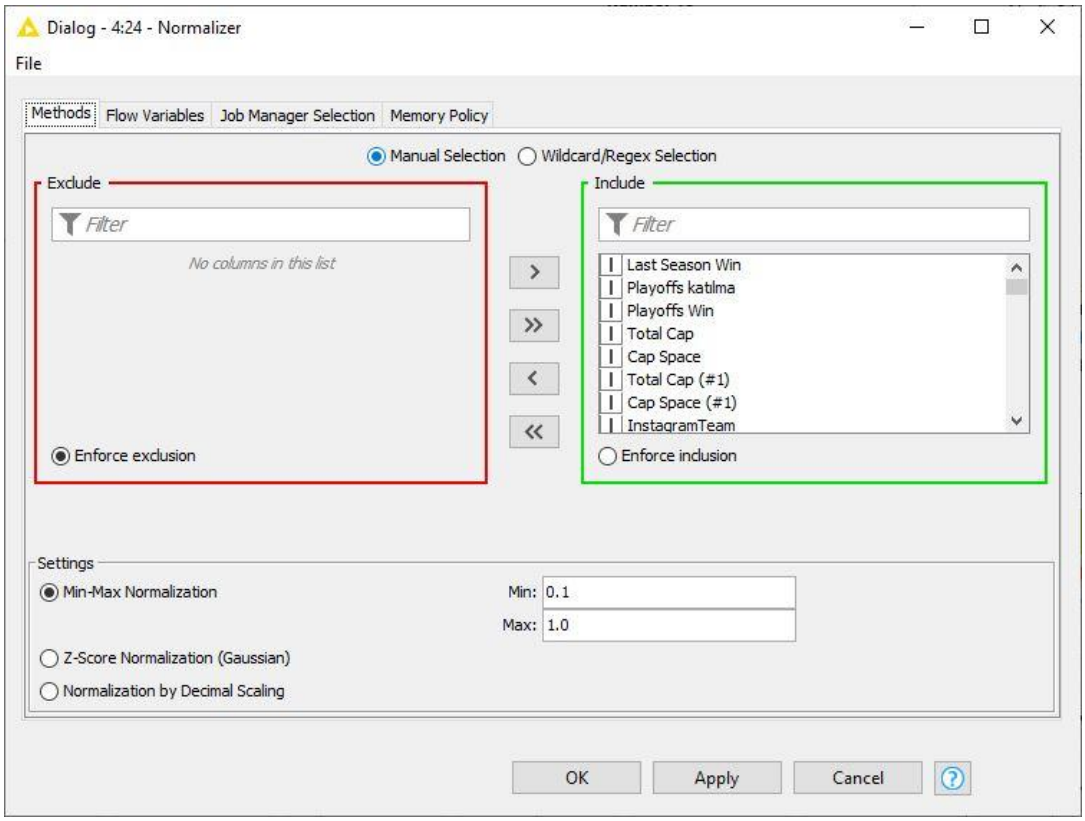
Veri setinde string olarak bulunan eski Pos değişkeninin olduğu sütunu modele dahil etmemek için tekrar Column Filter düğümü kullanılmış ve eski pos değişkeni veri setinden çıkartılmıştır. Şekil 4.21'de düğüm ayarları gösterilmiştir.



Şekil 4.21 Column Filter Düğümü

4.6.11. Normalizer Düğümü

Normalizer düğümü veri setindeki değerleri ölçeklendirme ve normalizasyon işlemlerini gerçekleştirmek için kullanılır. Bu düğüm veri setindeki değerleri istenen bir aralığa dönüştürmek amacıyla kullanılabilir. Standartlaştırma bazı makine öğrenmesi algoritmalarında daha iyi performans göstermesine yardımcı olabilmektedir. Yapay sinir ağları ve regresyon modellerini yaparken normalleştirme yapılması önemlidir. NBA veri seti farklı ölçek tiplerinde girdileri olan, çok büyük ve çok küçük verileri olan bir veri seti olduğu için bunları ortak bir ölçekte toplamak amacıyla normalize edilmiştir. Ayarlarda sayısal sütunlar modele katılmıştır. Şekil 4.22'de görüldüğü gibi Min-max yöntemi minimum değeri 0.1, maximum değeri 1 olacak şekilde seçilmiştir.



Şekil 4.22 Normalizer Düğümü

Şekil 4.23'de veri setinin normalize edilmiş hali yer almaktadır.

Normalized table - 4:24 - Normalizer

File Edit Hilite Navigation View

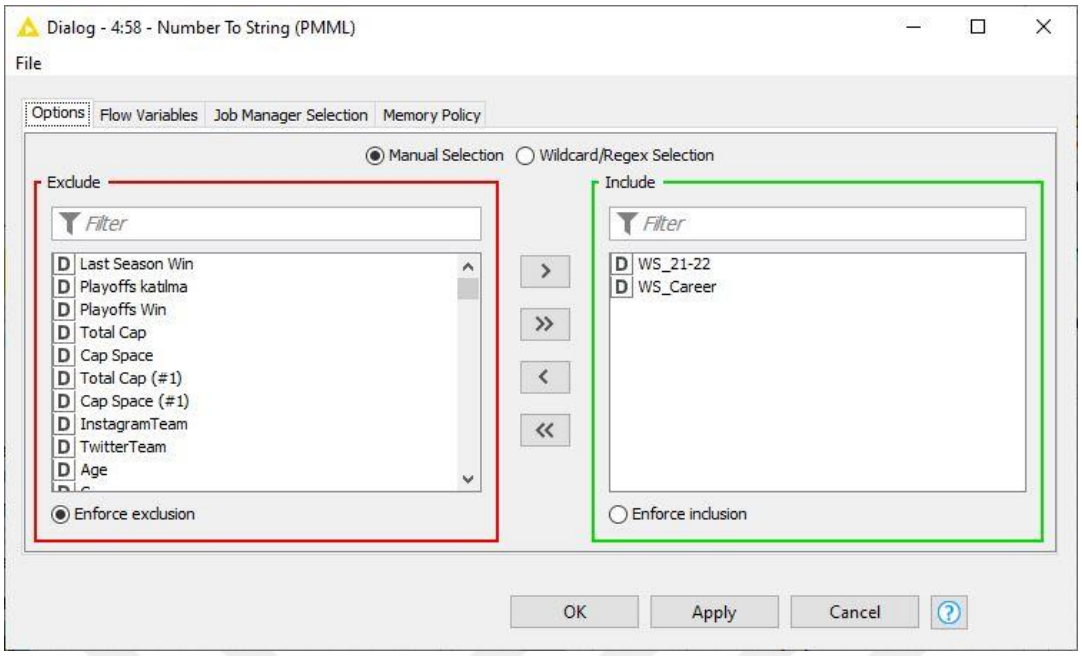
Table "default" - Rows: 240 Spec - Columns: 91 Properties Flow Variables

Row ID	Last Se...	Playoff...	Playoff...	Total Cap	Cap Sp...	Total C...	Cap Sp...	Instagr...	
Row0	1	1	0.494	0.516	0.584	0.777	0.323	0.159	0.14
Row1	1	1	0.494	0.516	0.584	0.777	0.323	0.159	0.14
Row2	1	1	0.494	0.516	0.584	0.777	0.323	0.159	0.14
Row3	1	1	0.494	0.516	0.584	0.777	0.323	0.159	0.14
Row4	1	1	0.494	0.516	0.584	0.777	0.323	0.159	0.14
Row5	1	1	0.494	0.516	0.584	0.777	0.323	0.159	0.14
Row6	1	1	0.494	0.516	0.584	0.777	0.323	0.159	0.14
Row7	1	1	0.494	0.516	0.584	0.777	0.323	0.159	0.14
Row8	0.836	1	0.438	0.345	0.755	0.336	0.764	0.116	0.13
Row9	0.836	1	0.438	0.345	0.755	0.336	0.764	0.116	0.13
Row10	0.836	1	0.438	0.345	0.755	0.336	0.764	0.116	0.13
Row11	0.836	1	0.438	0.345	0.755	0.336	0.764	0.116	0.13
Row12	0.836	1	0.438	0.345	0.755	0.336	0.764	0.116	0.13
Row13	0.836	1	0.438	0.345	0.755	0.336	0.764	0.116	0.13
Row14	0.836	1	0.438	0.345	0.755	0.336	0.764	0.116	0.13
Row15	0.836	1	0.438	0.345	0.755	0.336	0.764	0.116	0.13
Row16	0.775	1	1	1	0.1	1	0.1	1	0.72
Row17	0.775	1	1	1	0.1	1	0.1	1	0.72
Row18	0.775	1	1	1	0.1	1	0.1	1	0.72
Row19	0.775	1	1	1	0.1	1	0.1	1	0.72
Row20	0.775	1	1	1	0.1	1	0.1	1	0.72
Row21	0.775	1	1	1	0.1	1	0.1	1	0.72
Row22	0.775	1	1	1	0.1	1	0.1	1	0.72
Row23	0.775	1	1	1	0.1	1	0.1	1	0.72
Row24	0.775	1	0.719	0.565	0.535	0.567	0.533	0.258	0.44
Row25	0.775	1	0.719	0.565	0.535	0.567	0.533	0.258	0.44
Row26	0.775	1	0.719	0.565	0.535	0.567	0.533	0.258	0.44
Row27	0.775	1	0.719	0.565	0.535	0.567	0.533	0.258	0.44
Row28	0.775	1	0.719	0.565	0.535	0.567	0.533	0.258	0.44
Row29	0.775	1	0.719	0.565	0.535	0.567	0.533	0.258	0.44
Row30	0.775	1	0.719	0.565	0.535	0.567	0.533	0.258	0.44
Row31	0.775	1	0.719	0.565	0.535	0.567	0.533	0.258	0.44
Row32	0.755	1	0.606	0.641	0.459	0.814	0.286	0.172	0.19
Row33	0.755	1	0.606	0.641	0.459	0.814	0.286	0.172	0.19

Şekil 4.23 Normalize Veri Seti

4.6.12. Number to String (PMML)

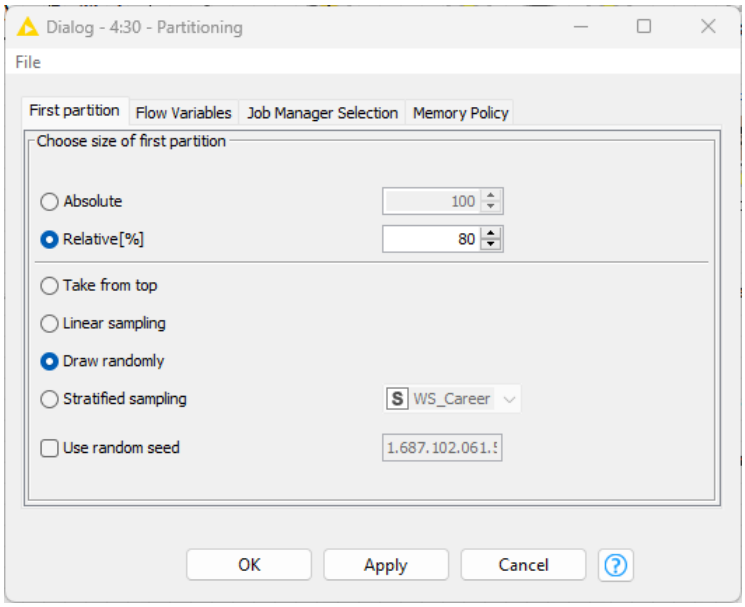
String tipiyle çalışan algoritmaları kullanabilmek için sayısal değerleri olan değişkenleri String değerlere çeviren Number to String düğümü kullanılmıştır. Şekil 4.24'de yer aldığı gibi çıktı değişkeni WS_Career ve WS_21-22 String tipine çevrilmiştir.



Şekil 4.24 Number to String Düğümü

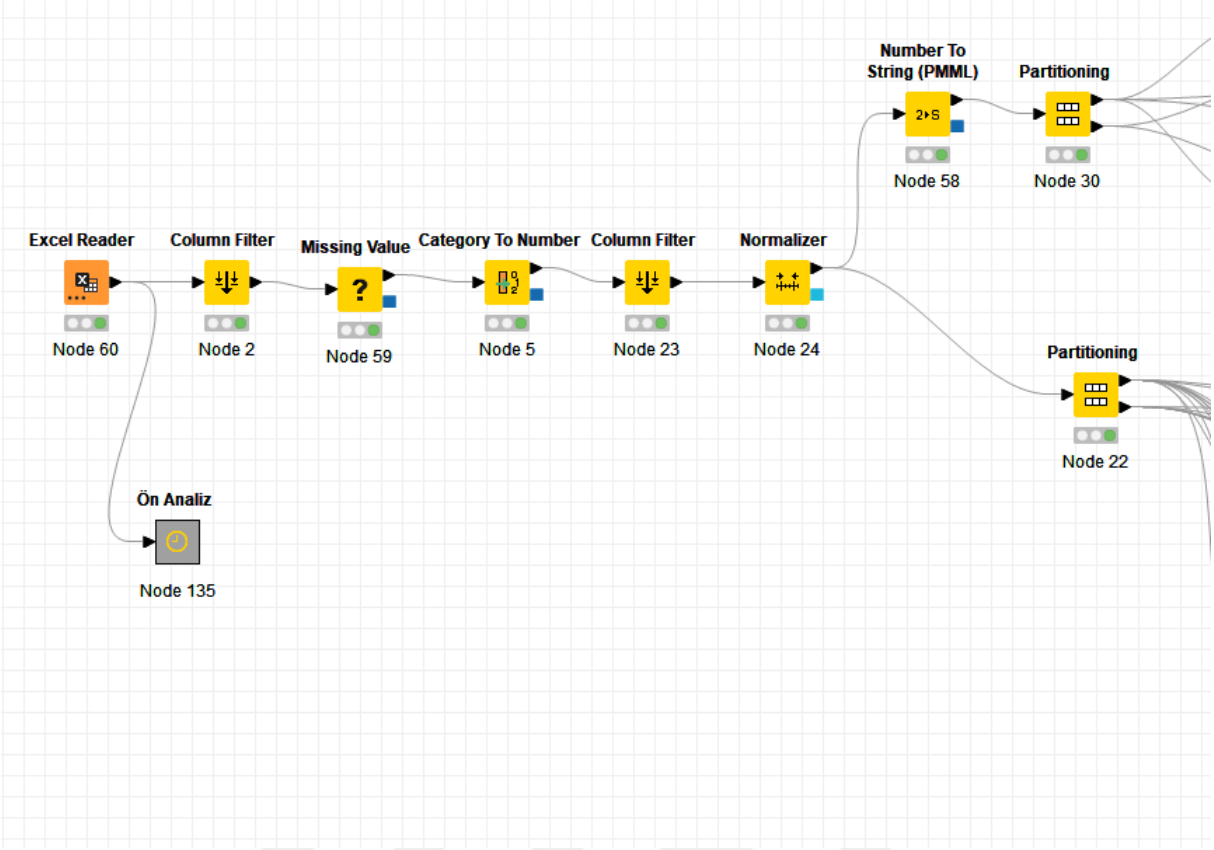
4.6.13. Partitioning

Veri setini eğitim ve test verisi olarak ikiye ayırmak için bu düğüm kullanılır. Bu düğümün ilk çıktısı eğitim datası, ikinci çıktısı test datası olarak kullanılmıştır. Veri setinin %80'i eğitim ve %20'si test verisi olmak üzere ikiye ayrılmıştır. Bu veride örnekleme yöntemi olarak rastgele örnekleme (Draw Randomly) kullanılmıştır (Şekil 4.25).



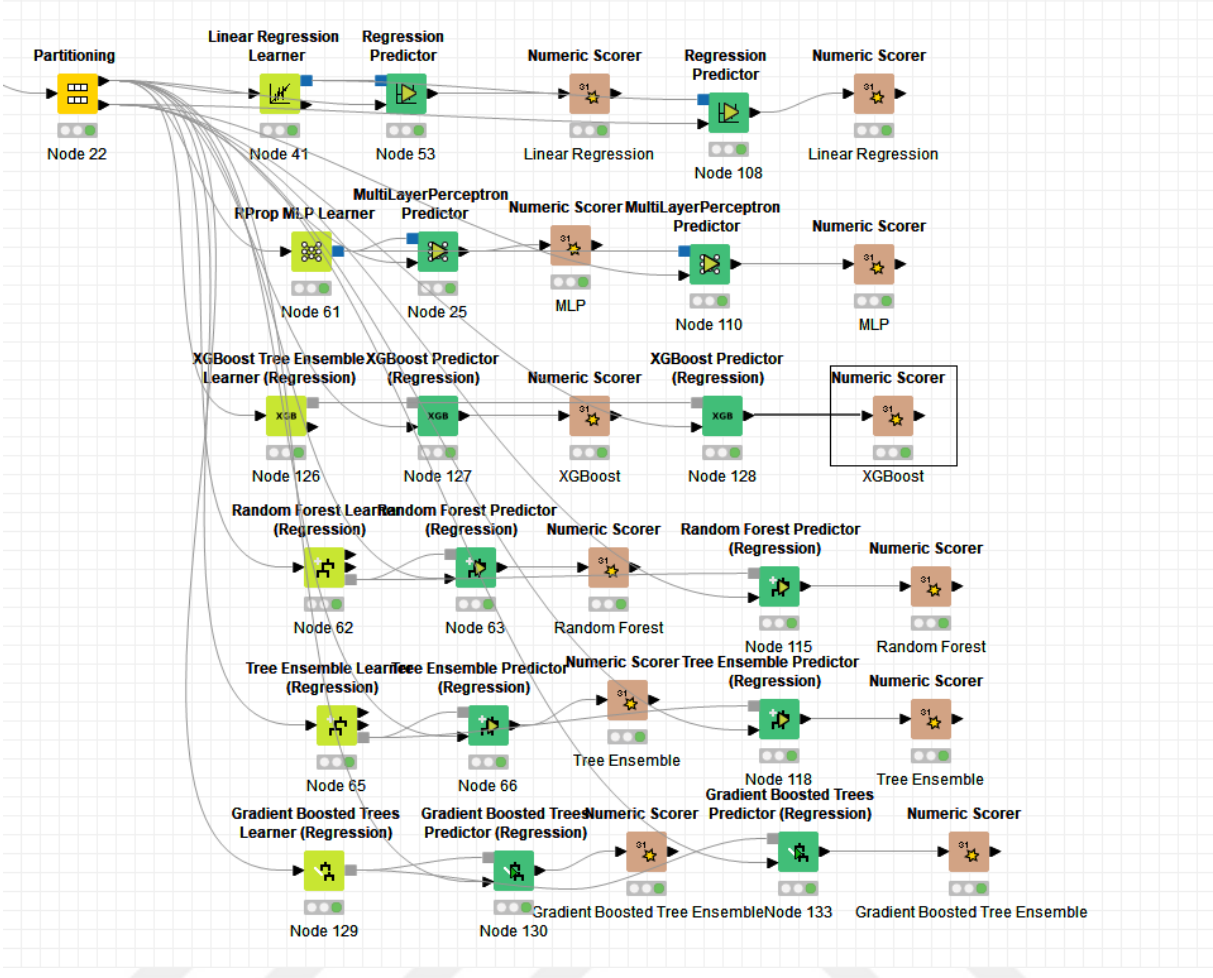
Şekil 4.25 Partitioning Düğümü

Uygulamada kullanılan düğümlerin Knime programındaki görünümü Şekil 4.26'de yer almaktadır.



Şekil 4.26 Knime Programında Görünüm

Çalışmada kullanılan datayı en iyi yansıtan makine öğrenmesi modelini seçmek için bazı modeller sırasıyla test edilmiştir. Lineer Regresyon, Yapay Sinir Ağları, XgBoost Learning, Karar Ağaçları, Karar Ağaçları Topluluk Öğrenmesi modellerinin Knime programındaki görünümü Şekil 4.27'da yer almaktadır.



Şekil 4.27 Modellerin Knime Programında Görünümü

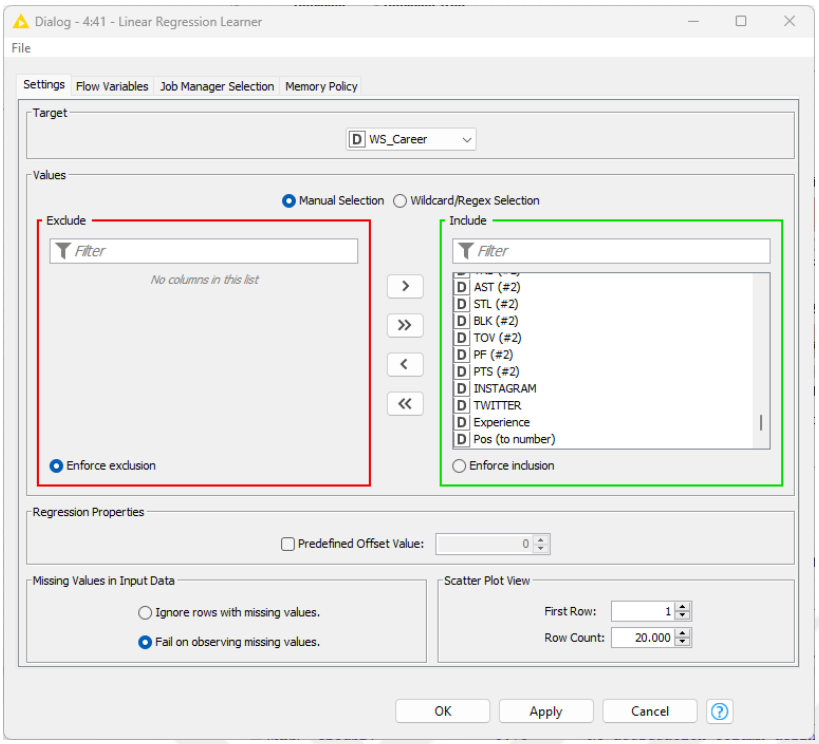
Knime modelleri öğrenme düğümü (Learner) ve tahmin düğümü (Predictor) olarak ikiye ayırır. Bütün modellerin iki tane düğümü vardır. Biri eğitim verisinin sonuçlarını, diğeri ise test verisinin sonuçlarını yansıtmaktadır.

4.6.14. Lineer Regresyon

Lineer Regresyon düğümü, veri setindeki bağımsız değişkenlerin hedef değişken üzerindeki etkisini incelemek için kullanılmıştır.

4.6.14.1. Linear Regression Learner

İlk aşama olan eğitim aşamasında modelin öğrenmesi Linear Regression Learner düğümüne Partitioning düğümünde ikiye ayrılan verinin %80'lik kısmı olan eğitim verisi ayarlanmıştır. Düğüm ayarlarında hedef değişken olarak WS_Career olarak seçilmiştir (Şekil 4.28).

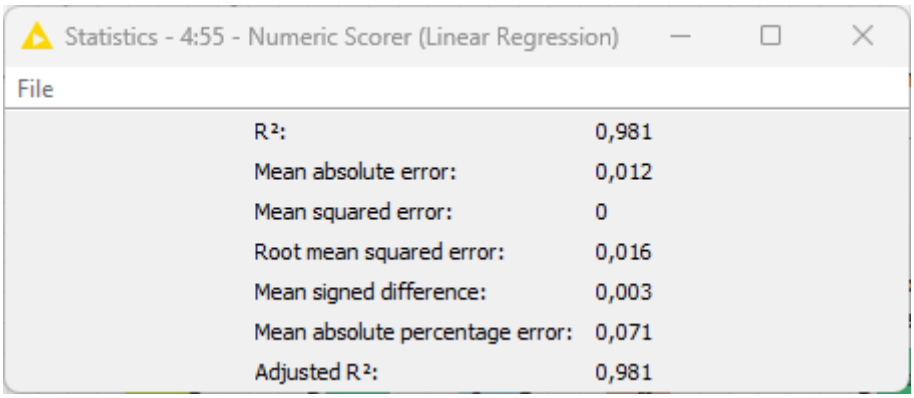


Şekil 4.28 Lineer Regresyon Düğümü

4.6.14.2. Regression Predictor ve Scorer -Eğitim

Tahmin düğümüne model bilgisi öğrenme düğümü olan Linear Regression Learner'den, tahmin edilecek veri bilgisi de Partitioning Düğümü'nde ayrılan tüm verinin %80'lik kısmı olan eğitim verisinden gelecek şekilde bağlanmış ve çalıştırılmıştır.

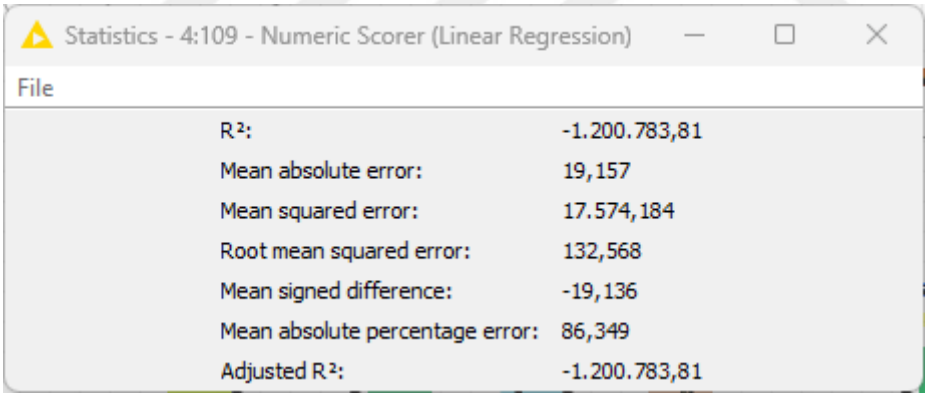
Regression Predictor düğümüne performans ölçümü için kullanılan Scorer düğümü bağlanmıştır. Ayarlarda gerçek değer olan "first column" "WS_Career", tahmin olan "second column" ise Prediction (WS_Career) olarak ayarlanmıştır. Scorer düğümü, bir modelin performansını değerlendirmek için kullanılan bir düğümdür. Daha düşük bir MAE, MSE veya RMSE değeri, daha iyi bir tahmin performansını gösterirken, daha yüksek bir R-kare değeri, modelin daha iyi bir uyum sağladığını gösterir. Şekil 4.29'de görüldüğü üzere scorer sonuçlarına bakıldığında R-Kare değeri 0,981, MAE değeri 0,012, MSE değeri 0, RMSE değeri 0,016, MAPE değeri 0,071 olarak bulunmuştur.



Şekil 4.29 Lineer Regresyon Eğitim Sonuçları

4.6.14.3. Regression Predictor ve Scorer -Test

Eğitim için olan Linear Regression Predictor ve Scorer düğümleri ayarlarıyla birlikte aynı şekilde uygulanmıştır. Tahmin düğümüne model bilgisi öğrenme düğümü olan Linear Regression Learner'den, tahmin edilecek veri bilgisi de Partitioning Düğümü'nde ayrılan tüm verinin %20'lik kısmı olan test verisinden gelecek şekilde bağlanmış ve çalıştırılmıştır. Şekil 4.29'da görülen scorer sonuçlarına bakıldığında R-Kare değeri-1.200.783,81, MAE değeri 19,157, MSE değeri 17,574,184, RMSE değeri 132,568, MAPE değeri 86,349 olarak bulunmuştur.



Şekil 4.30 Lineer Regresyon Test Sonuçları

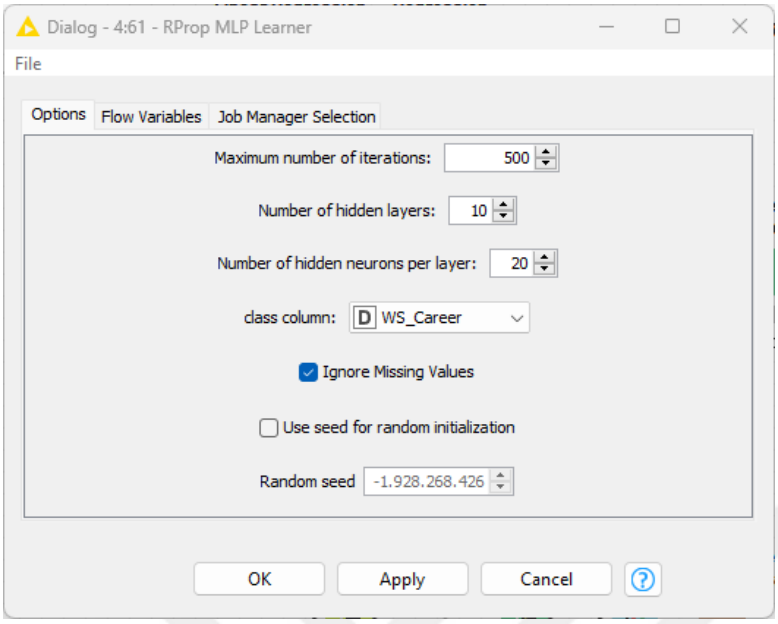
4.6.15. Yapay Sinir Ağları (MLP)

MLP düğümü yapay sinir ağındaki gizli katman sayısını ve her bir katmandaki düğüm sayısını ayarlamak için kullanılmıştır. Ayrıca aktivasyon fonksiyonlarını, iterasyon sayısını, öğrenme hızını ve diğer parametreleri de yapılandırılmıştır.

4.6.15.1. RProp MLP Learner

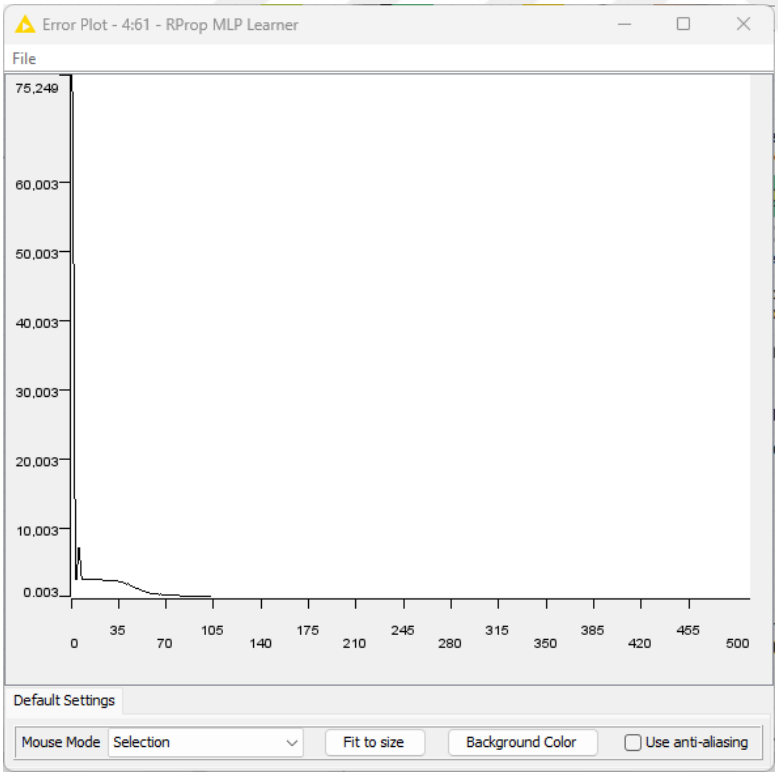
Modelin eğitim aşamasında RProp MLP Learner düğümüne Partitioning düğümünde ikiye ayrılan verinin %80'lik kısmı olan eğitim verisi bağlanmıştır. Ayarlarda "class column" yani

çıktı WS_Career olarak seçilmiş, iterasyon sayısı 500'e çıkarılmıştır. Katman sayısı 10, gizli nöron sayısı ise 20 olarak ayarlanmıştır (Şekil 4.31).



Şekil 4.31 MLP Learner Düğümü

Şekil 4.32 incelendiğinde eğitim aşamasında hata grafiğinde yakınsamanın sağlandığı görülmektedir.

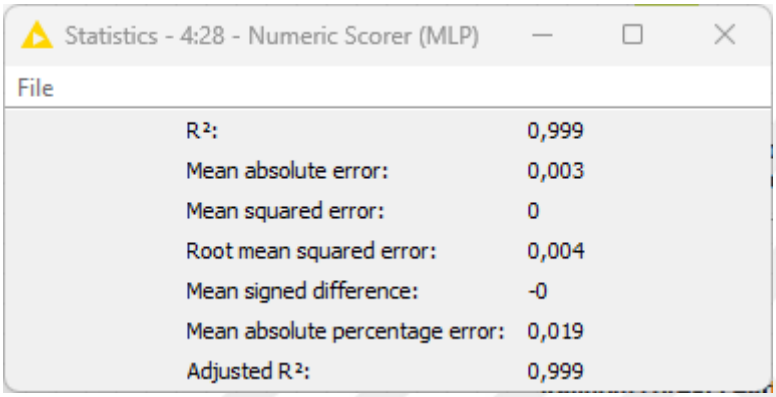


Şekil 4.32 MLP Learner

4.6.15.2. MLP Predictor ve Scorer -Eğitim

Tahmin düğümüne model bilgisi öğrenme düğümü olan MLP Learner'den, tahmin edilecek veri bilgisi de Partitioning Düğümü'nde ayrılan tüm verinin %80'lik kısmı olan eğitim verisinden gelecek şekilde bağlanmış ve çalıştırılmıştır.

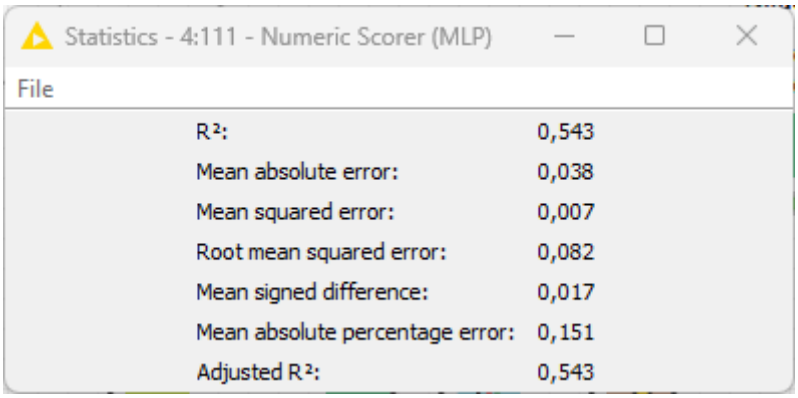
MLP Predictor düğümüne performans ölçümü için kullanılan Scorer düğümü bağlanmıştır. Ayarlarda gerçek değer olan "first column" "WS_Career", tahmin olan "second column" ise Prediction (WS_Career) olarak ayarlanmıştır. Scorer sonuçlarına bakıldığında R-Kare değeri 0,999, MAE değeri 0,003, MSE değeri 0, RMSE değeri 0,004, MAPE değeri 0,019 olarak bulunmuştur (Şekil 4.33).



Şekil 4.33 MLP Eğitim Sonuçları

4.6.15.3. MLP Predictor ve Scorer -Test

Eğitim için olan MLP Predictor ve Scorer düğümleri ayarlarıyla birlikte aynı şekilde uygulanmıştır. Tahmin düğümüne model bilgisi öğrenme düğümü olan MLP Learner'den, tahmin edilecek veri bilgisi de Partitioning Düğümü'nde ayrılan tüm verinin %20'lik kısmı olan test verisinden gelecek şekilde bağlanmış ve çalıştırılmıştır. Scorer sonuçlarına bakıldığında R-Kare değeri 0,543, MAE değeri 0,038, MSE değeri 0,007, RMSE değeri 0,082, MAPE değeri 0,151 olarak bulunmuştur (Şekil 4.34).

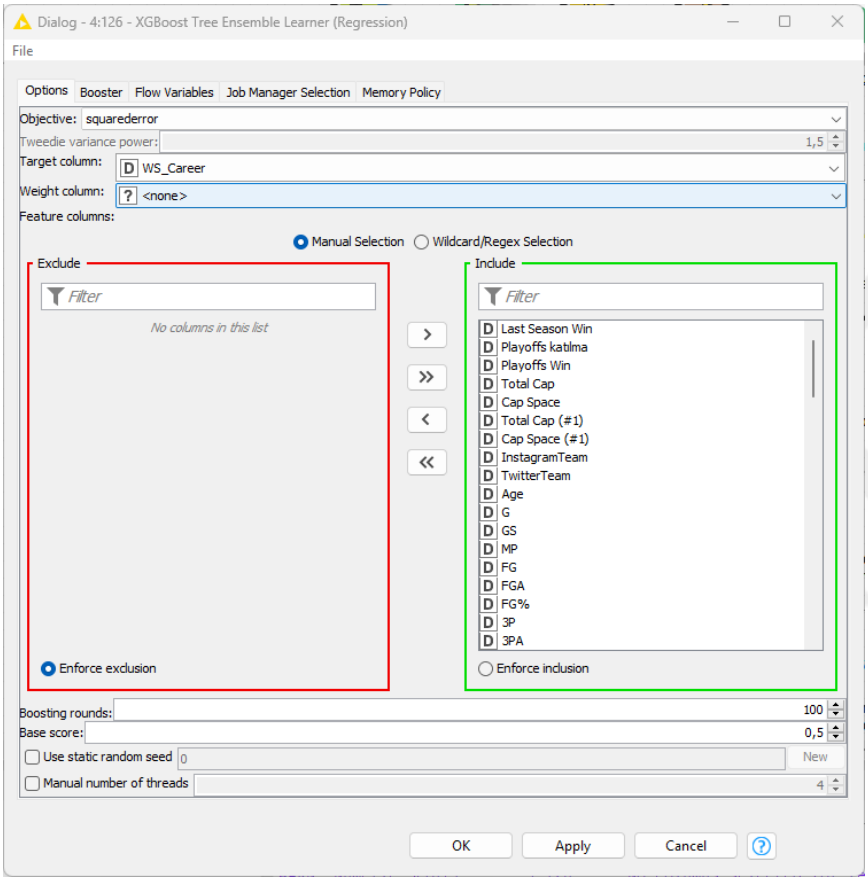


Şekil 4.34 MLP Test Sonuçları

4.6.16. XgBoost

4.6.16.1. XGBoost Tree Ensemble Learner

Modelin eğitim aşamasında benzer şekilde XGBoost Tree Ensemble Learner düğümüne Partitioning düğümünde ikiye ayrılan verinin %80'lik kısmı olan eğitim verisi bağlanmıştır. Ayarlar default olarak bırakıldı ve "target column" yani çıktı WS_Career olarak seçilmiştir (Şekil 4.35).

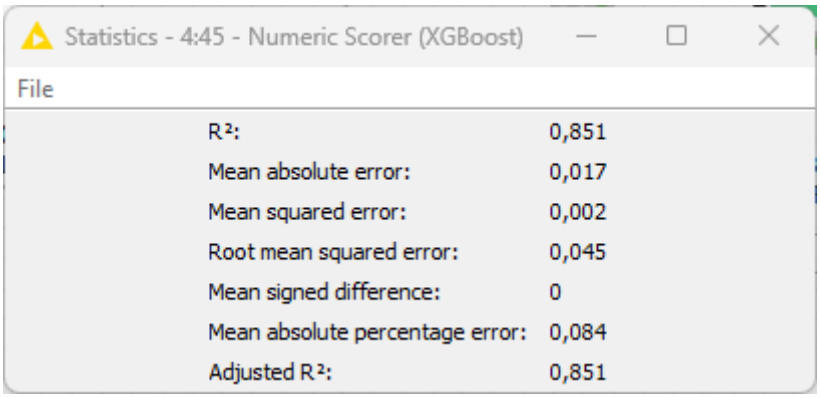


Şekil 4.35 XGBoost Düğümü

4.6.16.2. XGBoost Tree Ensemble Predictor ve Scorer -Eğitim

Tahmin düğümüne model bilgisi öğrenme düğümü olan XGBoost Tree Ensemble Learner'den, tahmin edilecek veri bilgisi de Partitioning Düğümü'nde ayrılan tüm verinin %80'lik kısmı olan eğitim verisinden gelecek şekilde bağlanmış ve çalıştırılmıştır.

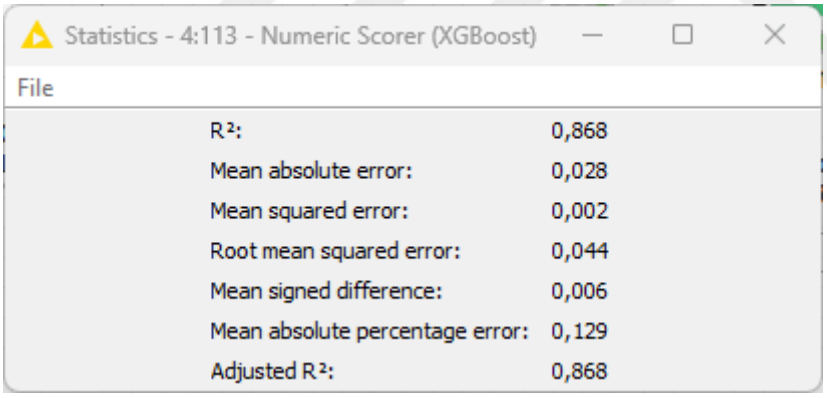
XGBoost Tree Ensemble Predictor düğümüne performans ölçümü için kullanılan Scorer düğümü bağlandı. Ayarlarda gerçek değer olan "first column" "WS_Career", tahmin olan "second column" ise Prediction (WS_Career) olarak ayarlandı. Şekil 4.36'de görülebileceği üzere scorer sonuçlarına bakıldığında R-Kare değeri 0,851, MAE değeri 0,017, MSE değeri 0,002, RMSE değeri 0,045, MAPE değeri 0,084 olarak bulunmuştur.



Şekil 4.36 XGBoost Düğümü Sonuçları Eğitim

4.6.16.3. XGBoost Tree Ensemble Predictor ve Scorer -Test

Eğitim için olan XGBoost Tree Ensemble Predictor ve Scorer düğümleri ayarlarıyla birlikte aynı şekilde uygulanmıştır. Tahmin düğümüne model bilgisi öğrenme düğümü olan XGBoost Tree Ensemble Learner'den, tahmin edilecek veri bilgisi de Partitioning Düğümü'nde ayrılan tüm verinin %20'lik kısmı olan test verisinden gelecek şekilde bağlanmış ve çalıştırılmıştır. Scorer sonuçlarına bakıldığında R-Kare değeri 0,868, MAE değeri 0,028, MSE değeri 0,002, RMSE değeri 0,044, MAPE değeri 0,129 olarak bulunmuştur (Şekil 4.37).

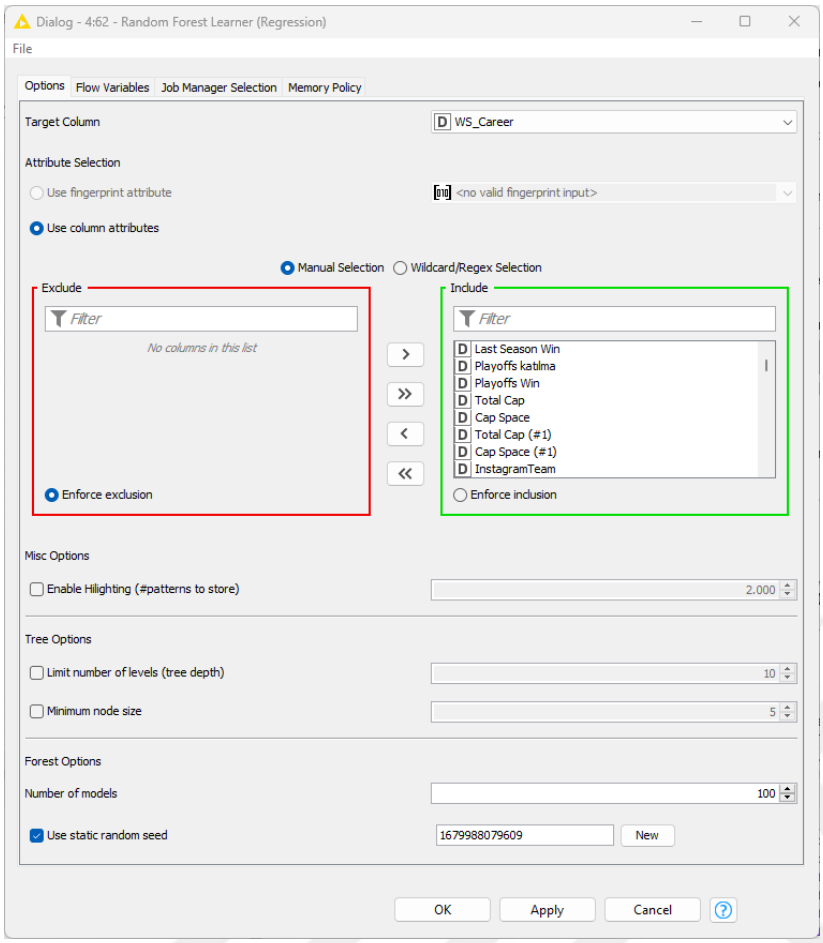


Şekil 4.37 XGBoost Düğümü Sonuçları Test

4.6.17. Rastgele Orman

4.6.17.1. Random Forest Learner (Regression)

Modelin eğitim aşamasında Random Forest Learner düğümüne Partitioning düğümünde ikiye ayrılan verinin %80'lik kısmı olan eğitim verisi bağlanmıştır. Ayarlar default olarak bırakılmış ve "target column" yani çıktı WS_Career olarak seçilmiştir (Şekil 4.38).

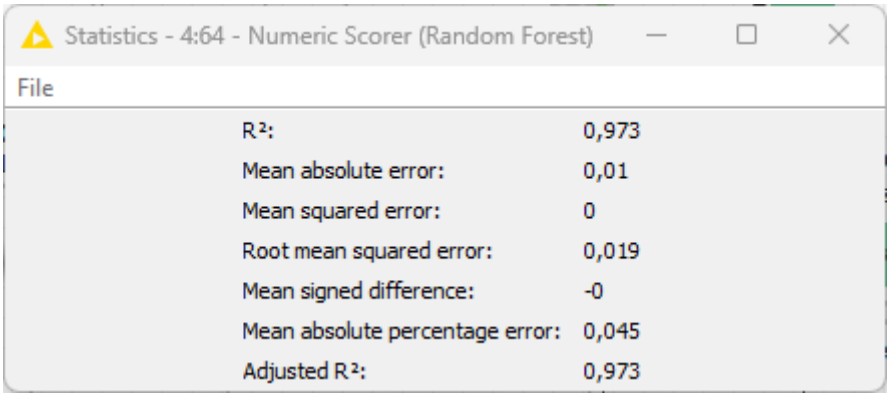


Şekil 4.38 Rastgele Orman Düğümü

4.6.17.2. Random Forest Predictor (Regression) ve Scorer -Eğitim

Tahmin düğümüne model bilgisi öğrenme düğümü olan Random Forest Learner'den, tahmin edilecek veri bilgisi de Partitioning Düğümü'nde ayrılan tüm verinin %80'lik kısmı olan eğitim verisinden gelecek şekilde bağlanmış ve çalıştırılmıştır.

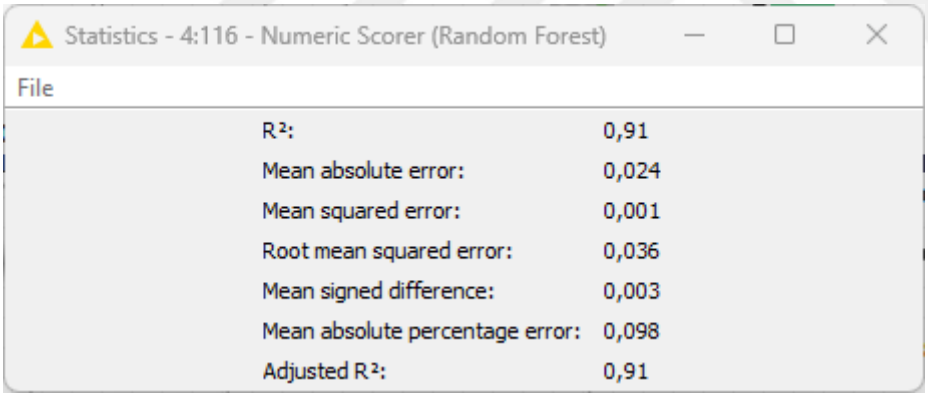
Random Forest Predictor düğümüne performans ölçümü için kullanılan Scorer düğümü bağlanmıştır. Ayarlarda gerçek değer olan "first column" "WS_Career", tahmin olan "second column" ise Prediction (WS_Career) olarak ayarlanmıştır. Scorer sonuçlarına bakıldığında R-Kare değeri 0,973, MAE değeri 0,01, MSE değeri 0, RMSE değeri 0,019, MAPE değeri 0,045 olarak bulunmuştur (Şekil 4.39).



Şekil 4.39 Rastgele Orman Sonuçları Eğitim

4.6.17.3. Random Forest Predictor (Regression) ve Scorer -Test

Eğitim için olan Random Forest Predictor ve Scorer düğümleri ayarlarıyla birlikte aynı şekilde uygulanmıştır. Tahmin düğümüne model bilgisi öğrenme düğümü olan Random Forest Learner'den, tahmin edilecek veri bilgisi de Partitioning Düğümü'nde ayrılan tüm verinin %20'lik kısmı olan test verisinden gelecek şekilde bağlanmış ve çalıştırılmıştır. Scorer sonuçlarına bakıldığında R-Kare değeri 0,91, MAE değeri 0,024, MSE değeri 0,001, RMSE değeri 0,036, MAPE değeri 0,098 olarak bulunmuştur (Şekil 4.40).



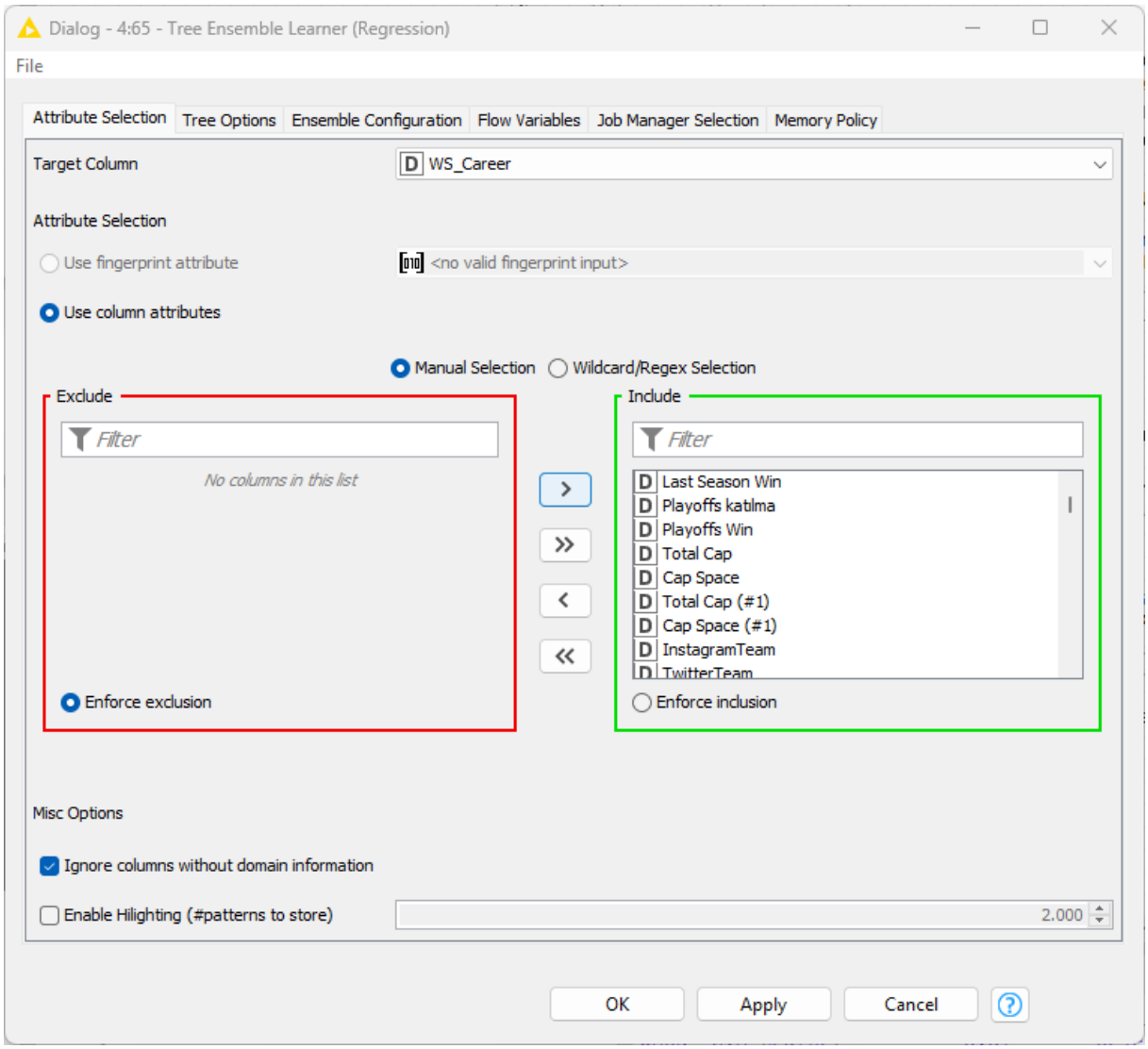
Şekil 4.40 Rastgele Orman Sonuçları Test

Knime programında Karar Ağaçları Topluluk Öğrenmesi altında Rastgele Orman ve Gradient Boosting olarak iki tane regresyon modeli bulunmaktadır. Bu modeller de veri setimiz üzerinde test edilmiştir.

4.6.18. Karar Ağaçları Topluluk Öğrenmesi Rastgele Orman

4.6.18.1. Tree Ensemble Learner (Regression)

Modelin eğitim aşamasında Tree Ensemble Learner düğümüne Partitioning düğümünde ikiye ayrılan verinin %80'lik kısmı olan eğitim verisi bağlanmıştır. Ayarlar default olarak bırakılmış ve "target column" yani çıktı WS_Career olarak seçilmiştir (Şekil 4.41).

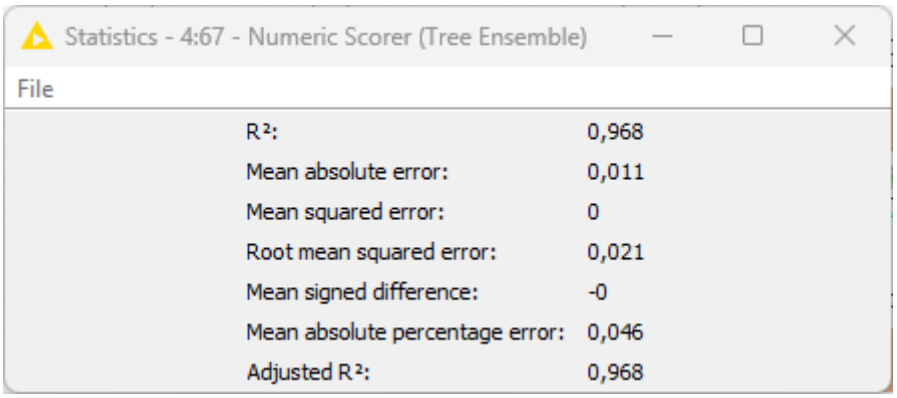


Şekil 4.41 Topluluk Öğrenmesi Düğümü

4.6.18.2. Tree Ensemble Predictor (Regression) ve Scorer -Eğitim

Tahmin düğümüne model bilgisi öğrenme düğümü olan Tree Ensemble Learner'den, tahmin edilecek veri bilgisi de Partitioning Düğümü'nde ayrılan tüm verinin %80'lik kısmı olan eğitim verisinden gelecek şekilde bağlanmış ve çalıştırılmıştır.

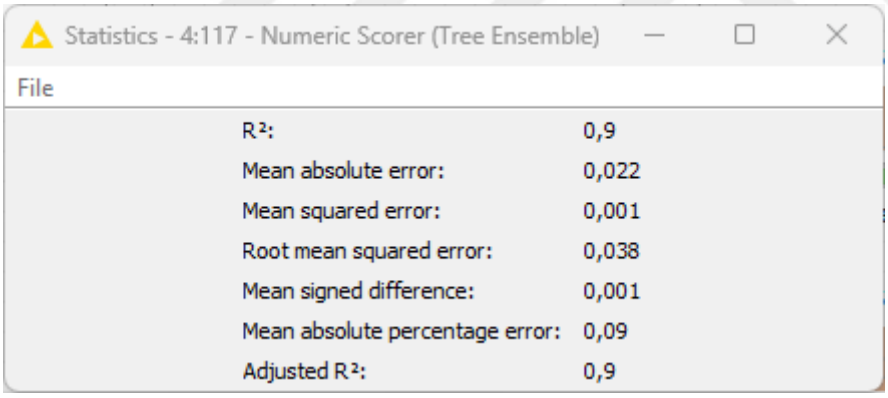
Tree Ensemble Predictor düğümüne performans ölçümü için kullanılan Scorer düğümü bağlanmıştır. Ayarlarda gerçek değer olan "first column" "WS_Career", tahmin olan "second column" ise Prediction (WS_Career) olarak ayarlanmıştır. Şekil 4.41'de de görüleceği gibi scorer sonuçlarına bakıldığında R-Kare değeri 0,968, MAE değeri 0,011, MSE değeri 0, RMSE değeri 0,021, MAPE değeri 0,046 olarak bulunmuştur.



Şekil 4.42 Topluluk Öğrenmesi Eğitim Sonuçları

4.6.18.3. Tree Ensemble Predictor (Regression) ve Scorer -Test

Eğitim için olan Tree Ensemble Predictor ve Scorer düğümleri ayarlarıyla birlikte aynı şekilde uygulanmıştır. Tahmin düğümüne model bilgisi öğrenme düğümü olan Tree Ensemble Learner'den, tahmin edilecek veri bilgisi de Partitioning Düğümü'nde ayrılan tüm verinin %20'lik kısmı olan test verisinden gelecek şekilde bağlanmış ve çalıştırılmıştır. Scorer sonuçlarına bakıldığında R-Kare değeri 0,9, MAE değeri 0,022, MSE değeri 0,001, RMSE değeri 0,038, MAPE değeri 0,09 olarak bulunmuştur (Şekil 4.43).

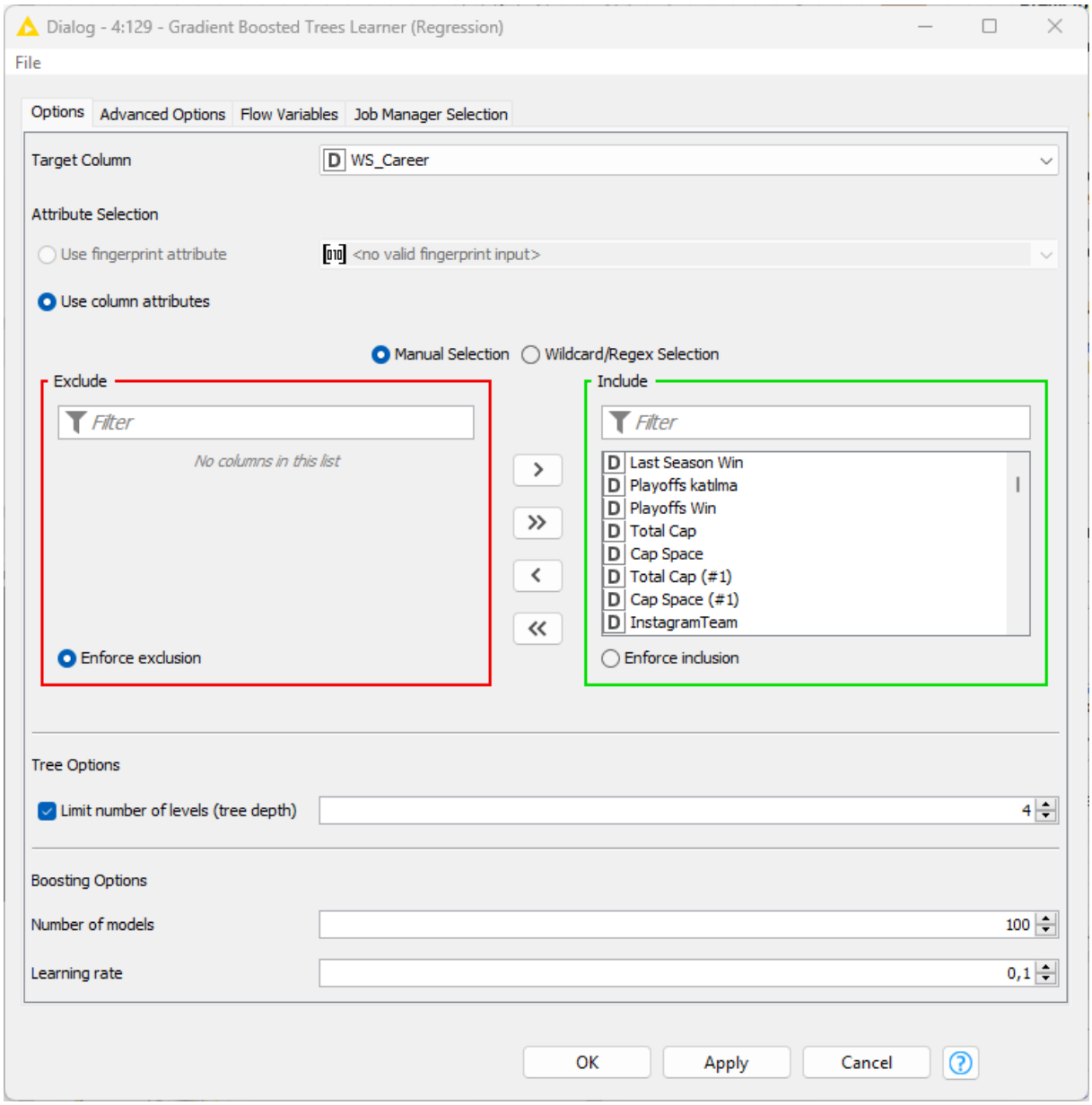


Şekil 4. 43 Topluluk Öğrenmesi Test Sonuçları

4.6.19. Karar Ağaçları Topluluk Öğrenmesi Gradient Boosting

4.6.19.1. Gradient Boosted Trees Learner (Regression)

Modelin eğitim aşamasında Gradient Boosted Trees Learner düğümüne Partitioning düğümünde ikiye ayrılan verinin %80'lik kısmı olan eğitim verisi bağlanmıştır. Ayarlar default olarak bırakılmış ve "target column" yani çıktı WS_Career olarak seçilmiştir (Şekil 4.44).

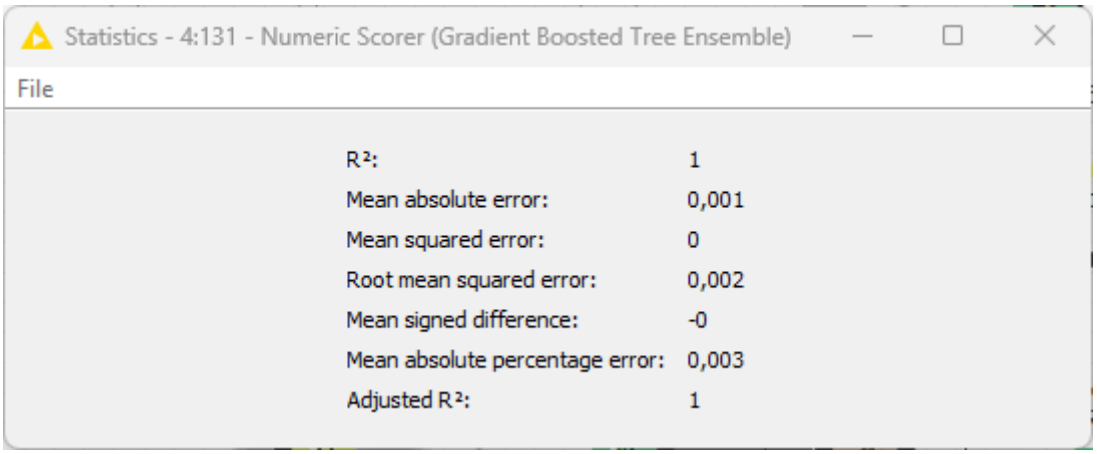


Şekil 4.44 Gradient Boosted Trees Düğümü

4.6.19.2. Gradient Boosted Trees Predictor (Regression) ve Scorer -Eğitim

Tahmin düğümüne model bilgisi öğrenme düğümü olan Gradient Boosted Trees Learner'den, tahmin edilecek veri bilgisi de Partitioning Düğümü'nde ayrılan tüm verinin %80'lik kısmı olan eğitim verisinden gelecek şekilde bağlanmış ve çalıştırılmıştır.

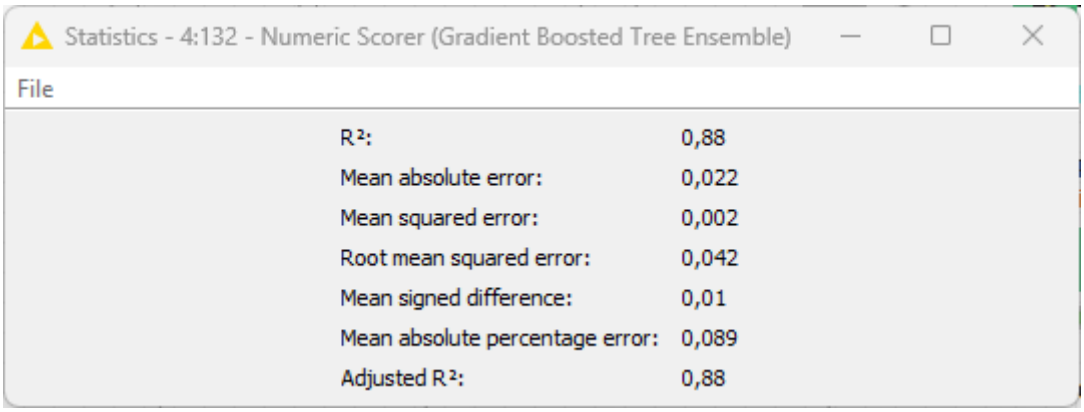
Gradient Boosted Trees Predictor düğümüne performans ölçümü için kullanılan Scorer düğümü bağlanmıştır. Ayarlarda gerçek değer olan "first column" "WS_Career", tahmin olan "second column" ise Prediction (WS_Career) olarak ayarlanmıştır. Scorer sonuçlarına bakıldığında R-Kare değeri 1, MAE değeri 0,001, MSE değeri 0, RMSE değeri 0,002, MAPE değeri 0,003 olarak bulunmuştur (Şekil 4.45).



Şekil 4.45 Gradient Boosted Trees Eğitim Sonuçları

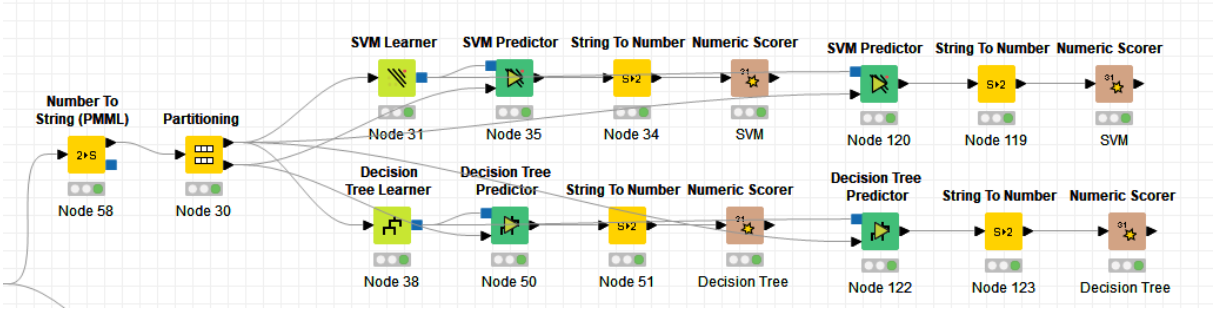
4.6.19.3. Gradient Boosted Trees Predictor (Regression) ve Scorer -Test

Eğitim için olan Gradient Boosted Trees Predictor ve Scorer düğümleri ayarlarıyla birlikte aynı şekilde uygulanmıştır. Tahmin düğümüne model bilgisi öğrenme düğümü olan Gradient Boosted Trees Learner'den, tahmin edilecek veri bilgisi de Partitioning Düğümü'nde ayrılan tüm verinin %20'lik kısmı olan test verisinden gelecek şekilde bağlanmış ve çalıştırılmıştır. Şekil 4.46'de gösteriliği gibi scorer sonuçlarına bakıldığında R-Kare değeri 0,88, MAE değeri 0,022, MSE değeri 0,002, RMSE değeri 0,042, MAPE değeri 0,089 olarak bulunmuştur.



Şekil 4.46 Gradient Boosted Trees Test Sonuçları

Bu aşamadan sonra anlatılacak SVM ve Karar Ağacı Modellerinin Knime Programında görünümü Şekil 4.47'da yer almaktadır.



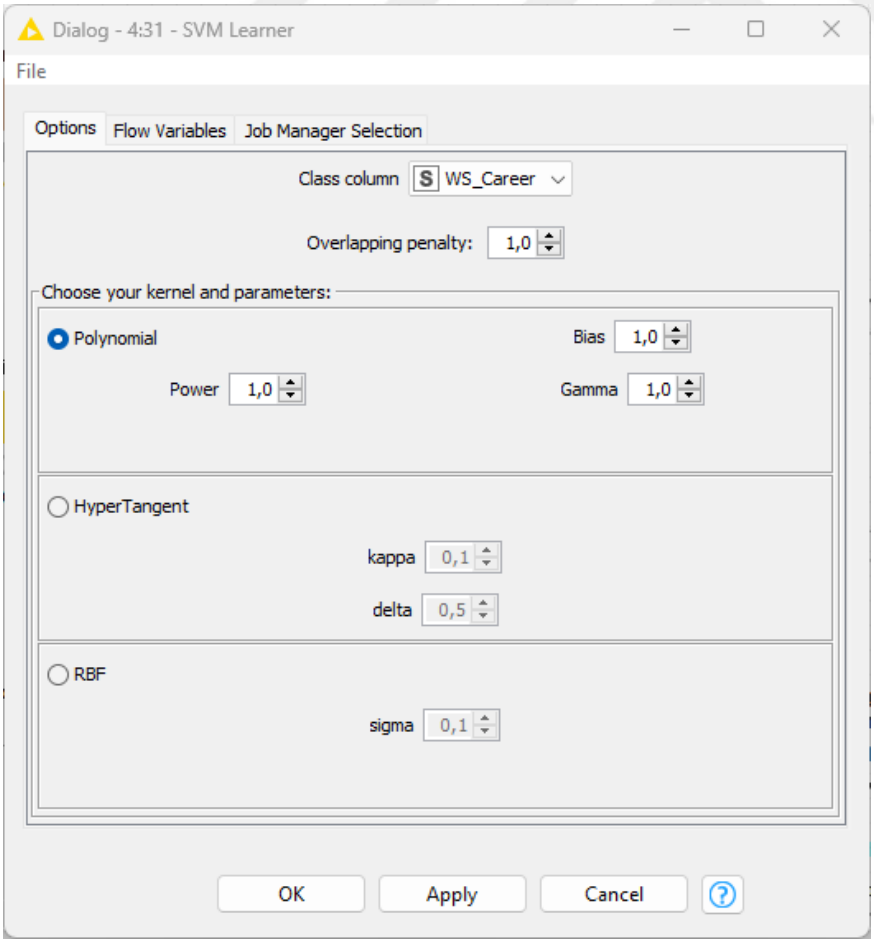
Şekil 4.47 SVM ve Karar Ağacı Modellerinin Knime Programında Görünümü

4.6.20. Destek Vektör Makineleri (SVM)

SVM modeli metinsel bir çıktı değeri beklediği için Number to String düğümü kullanarak Stringe çevirdiğimiz çıktı değişkeni modelde kullanılmıştır.

4.6.20.1. SVM Learner

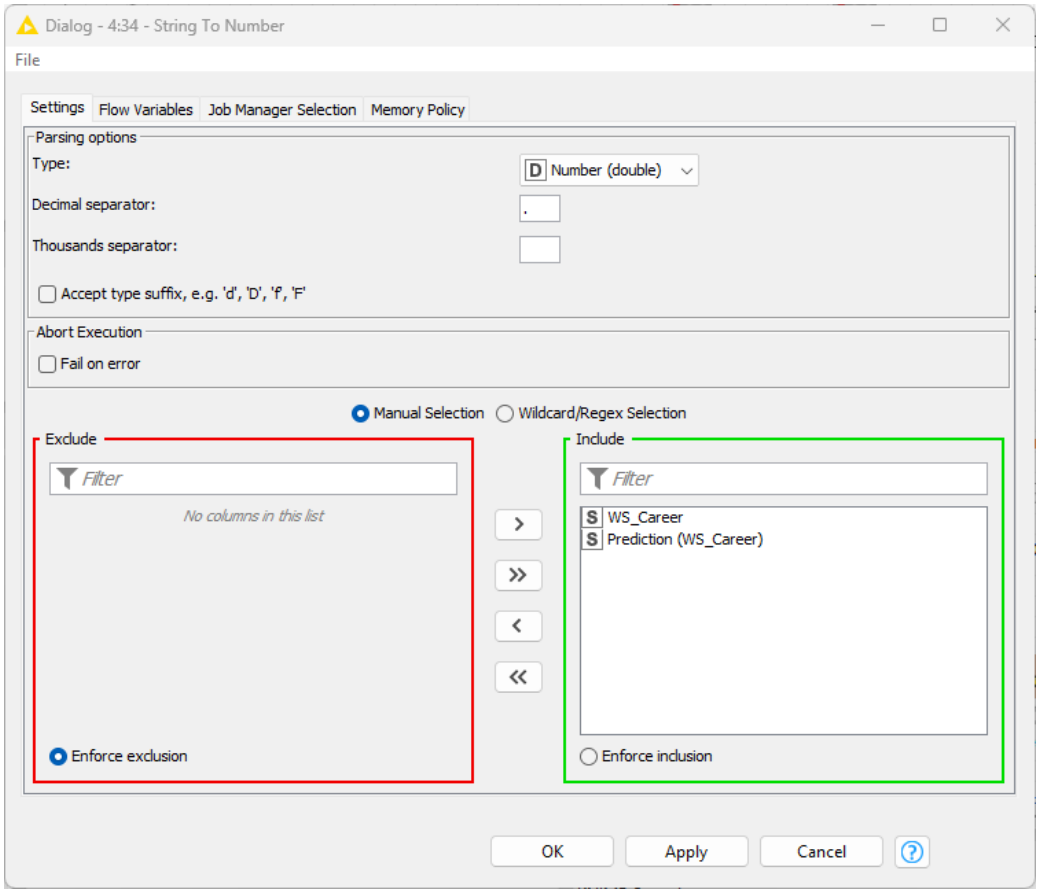
Modelin eğitim aşamasında SVM Learner düğümüne Partitioning düğümünde ikiye ayrılan verinin %80'lik kısmı olan eğitim verisi bağlanmıştır. Ayarlar default olarak bırakılmış ve "class column" yani çıktı WS_Career olarak seçilmiştir (Şekil 4.48).



Şekil 4.48 SVM Learner Düğümü

4.6.20.2. String to Number Düğümü

String olarak bulunan gerçek WS_Career çıktı değişkeni ile tahmin edilen WS_Career çıktı değişkeni sayısal değerlere tekrar bu düğüm ile çevrilmiştir (Şekil 4.49).

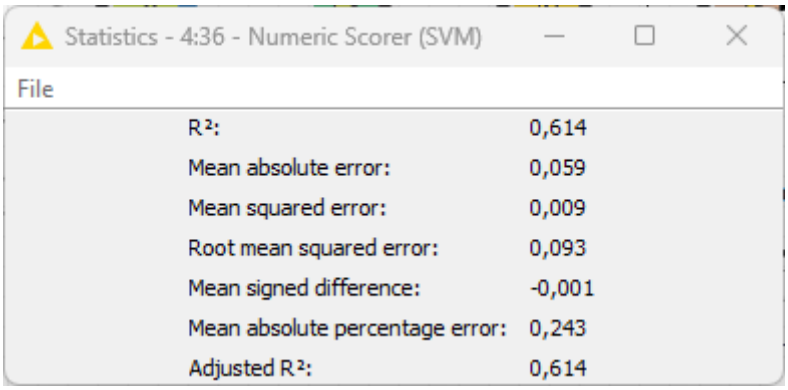


Şekil 4.49 String to Number Düğümü

4.6.20.3. SVM Predictor ve Scorer -Eğitim

Tahmin düğümüne model bilgisi öğrenme düğümü olan SVM Learner'den, tahmin edilecek veri bilgisi de Partitioning Düğümü'nde ayrılan tüm verinin %80'lik kısmı olan eğitim verisinden gelecek şekilde bağlanmış ve çalıştırılmıştır.

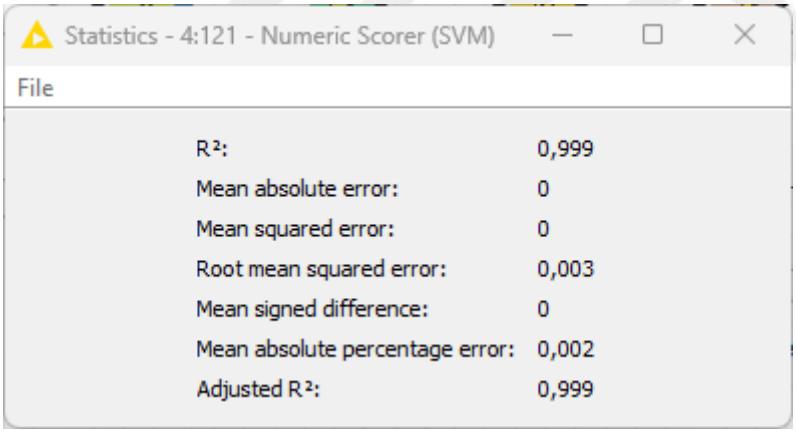
SVM Predictor düğümüne önce String to Number Düğümü bu düğümün sonucuna ise performans ölçümü için kullanılan Scorer düğümü bağlanmıştır. Ayarlarda gerçek değer olan "first column" "WS_Career", tahmin olan "second column" ise Prediction (WS_Career) olarak ayarlanmıştır. Scorer sonuçlarına bakıldığında R-Kare değeri 0,614, MAE değeri 0,059, MSE değeri 0,09, RMSE değeri 0,093, MAPE değeri 0,243 olarak bulunmuştur (Şekil 4.50).



Şekil 4.50 SVM Eğitim Sonuçları

4.6.20.4. SVM Predictor ve Scorer -Test

Eğitim için kullanılan SVM Predictor, Number to String ve Scorer düğümleri ayarlarıyla birlikte test ayarları da aynı şekilde uygulanmıştır. Tahmin düğümüne model bilgisi öğrenme düğümü olan SVM Learner'den, tahmin edilecek veri bilgisi de Partitioning Düğümü'nde ayrılan tüm verinin %20'lik kısmı olan test verisinden gelecek şekilde bağlanmış ve çalıştırılmıştır. Scorer sonuçlarına bakıldığında R-Kare değeri 0,999, MAE değeri 0, MSE değeri 0, RMSE değeri 0,003, MAPE değeri 0,002 olarak bulunmuştur (Şekil 4.51).



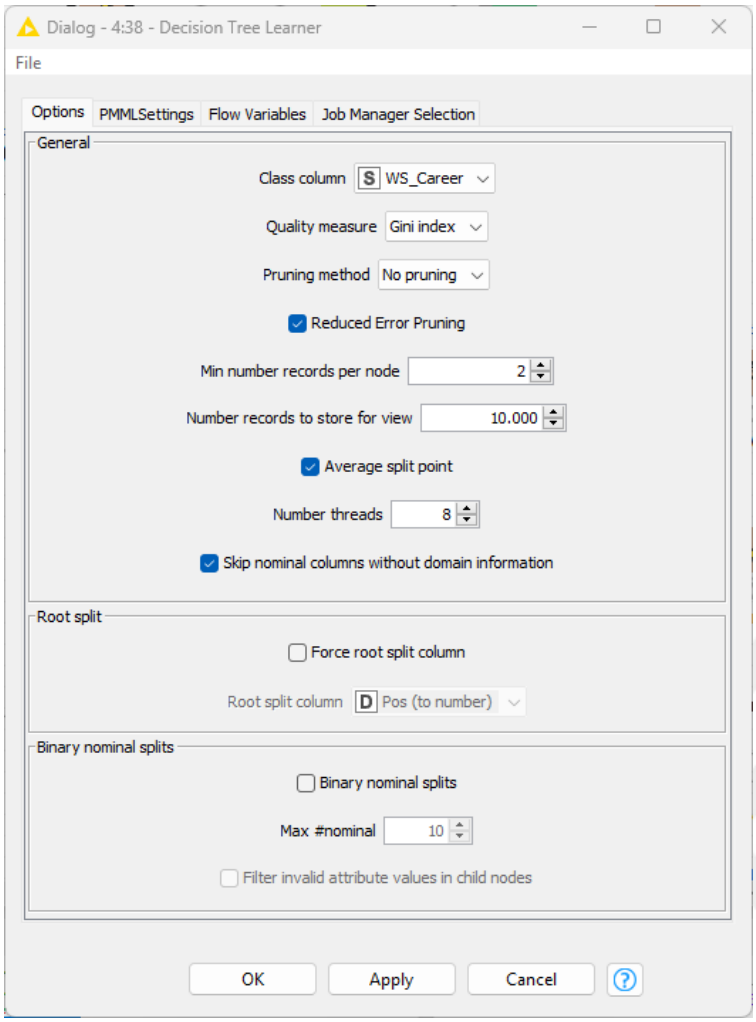
Şekil 4.51 SVM Test Sonuçları

4.6.21. Karar Ağacı

Karar Ağacı modeli metinsel bir çıktı değeri beklediği için Number to String düğümü kullanarak Stringe çevirdiğimiz çıktı değişkeni modelde kullanılmıştır.

4.6.21.1. Decision Tree Learner

Modelin eğitim aşamasında Decision Tree Learner düğümüne Partitioning düğümünde ikiye ayrılan verinin %80'lik kısmı olan eğitim verisi bağlanmıştır. Ayarlar default olarak bırakılmış ve "class column" yani çıktı WS_Career olarak seçilmiş, kalite "gini", budama "no puring", her düğümdeki örnek sayısı 2, ağacın derinliği (number threads) 8 olarak ayarlanmıştır (Şekil 4.52).



Şekil 4.52 Karar Ağacı Düğümü

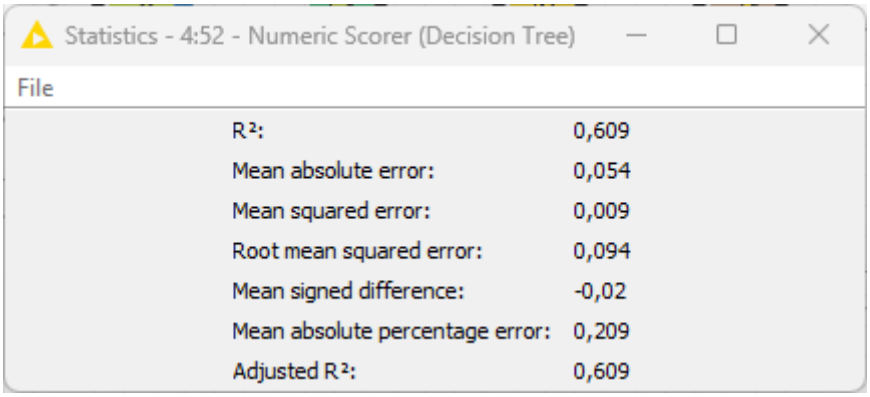
4.6.21.2. String to Number Düğümü

String olarak bulunan gerçek WS_Career çıktı değişkeni ile tahmin edilen WS_Career çıktı değişkeni sayısal değerlere tekrar bu düğüm ile çevrilmiştir.

4.6.21.3. Decision Tree Predictor ve Scorer -Eğitim

Tahmin düğümüne model bilgisi öğrenme düğümü olan Decision Tree Learner'den, tahmin edilecek veri bilgisi de Partitioning Düğümü'nde ayrılan tüm verinin %80'lik kısmı olan eğitim verisinden gelecek şekilde bağlanmış ve çalıştırılmıştır.

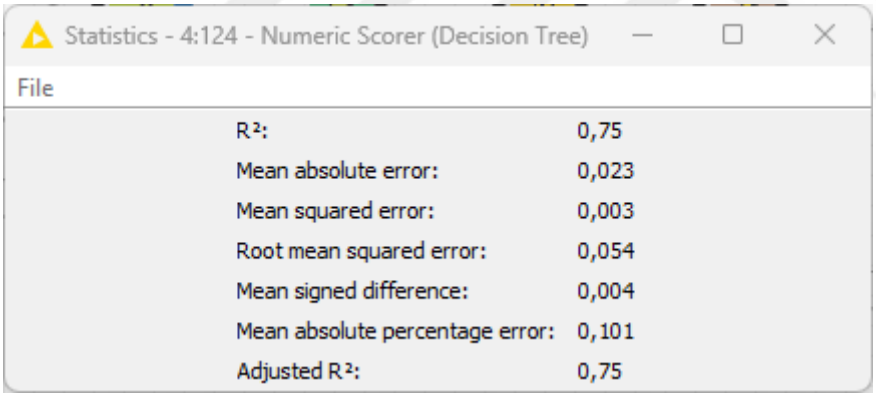
Decision Tree Predictor düğümüne önce String to Number Düğümü bu düğümün sonucuna ise performans ölçümü için kullanılan Scorer düğümü bağlanmıştır. Ayarlarda gerçek değer olan "first column" "WS_Career", tahmin olan "second column" ise Prediction (WS_Career) olarak ayarlanmıştır. Şekil 4.53'de gösterildiği gibi scorer sonuçlarına bakıldığında R-Kare değeri 0,609, MAE değeri 0,054, MSE değeri 0,009, RMSE değeri 0,094, MAPE değeri 0,209 olarak bulunmuştur.



Şekil 4.53 Karar Ağacı Eğitim Sonuçları

4.6.21.4. Decision Tree Predictor ve Scorer -Test

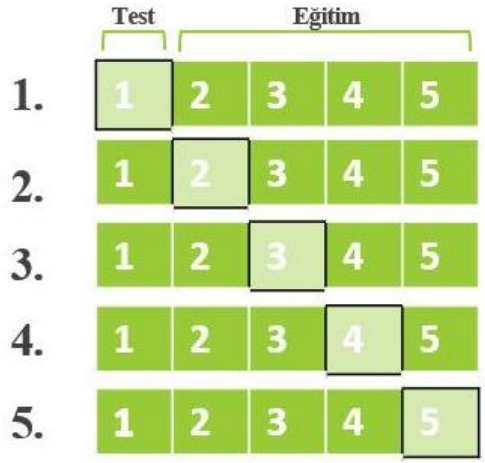
Eğitim için kullanılan SVM Predictor, Number to String ve Scorer düğümleri ayarlarıyla birlikte test ayarları aynı şekilde uygulanmıştır. Tahmin düğümüne model bilgisi öğrenme düğümü olan SVM Learner'den, tahmin edilecek veri bilgisi de Partitioning Düğümü'nde ayrılan tüm verinin %20'lik kısmı olan test verisinden gelecek şekilde bağlanmış ve çalıştırılmıştır. Scorer sonuçlarına bakıldığında R-Kare değeri 0,75, MAE değeri 0,023, MSE değeri 0,03, RMSE değeri 0,054, MAPE değeri 0,101 olarak bulunmuştur (Şekil 4.54).



Şekil 4.54 Karar Ağacı Test Sonuçları

4.6.22. Model Doğrulama- Çapraz Doğrulama (Cross Validation)

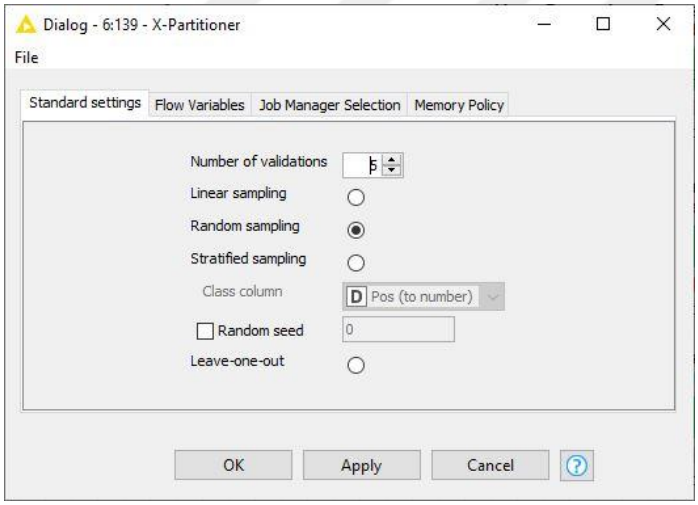
Modeli doğrulamak için modellerin ayarları değiştirilmeden X-Partitioner ve X-Aggregator düğümlerine bağlanarak Çapraz Doğrulama yapılmıştır. Çapraz Doğrulama, bir makine öğrenmesi modelinin performansını değerlendirmek için kullanılan bir yöntemdir ve veri setini eğitim ve test kümelerine bölerken, aynı veri seti üzerinde birden fazla kez tekrarlanır (Şekil 4.55). Bu sayede kısıtlı veri kümesine sahip çalışmaların güvenirlik testleri sağlanmış olur (Bergstra ve Bengio, 2012).



Şekil 4. 55 Çapraz Doğrulama

4.6.22.1. X-Partitioner

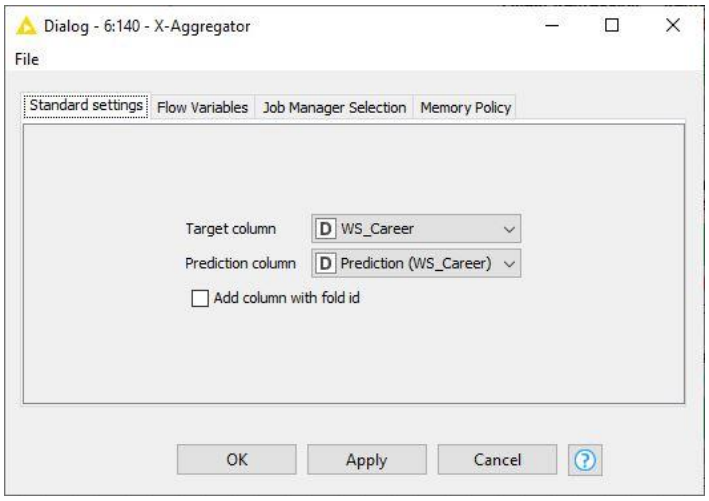
Knime'da X-Partitioner düğümü veri setini belirli bir kritere göre bölmek için kullanılmıştır. Veri setini benzer özelliklere sahip gruplara ayırmak, parçalara bölmek veya sınıflara göre gruplamak gibi farklı yöntemlerle ayırarak en iyi parametre ayarı yapılmaya çalışılmıştır. Şekil 4.56'de Yapay Sinir Ağları için X-Partitioner düğümünün ayarları yer almaktadır. Veri seti rastgele örnekleme ile 5 kümeye bölünmüştür.



Şekil 4.56 X-Partitioner

4.6.22.2. X-Aggregator

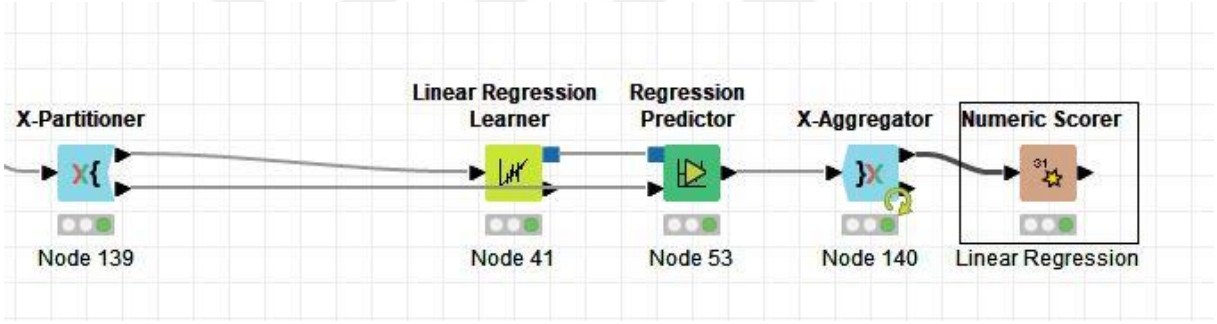
X-Partitioner düğümünden elde edilen parçalara ait sonuçları birleştirmek için X-Aggregator düğümü kullanılmıştır. X-Aggregator düğümünün ayarlarında gerçek değer olan "target column" "WS_Career", tahmin olan "prediction column" ise Prediction (WS_Career) olarak ayarlanmıştır (Şekil 4.57).



Şekil 4.57 X-Aggregator Düğümü

X-Partitioner ve X-Aggregator düğümlerinin Lineer Regresyon modeline bağlanması Şekil 4.58'da gösterilmiştir.

4.6.22.3. Lineer Regresyon Çapraz Doğrulama

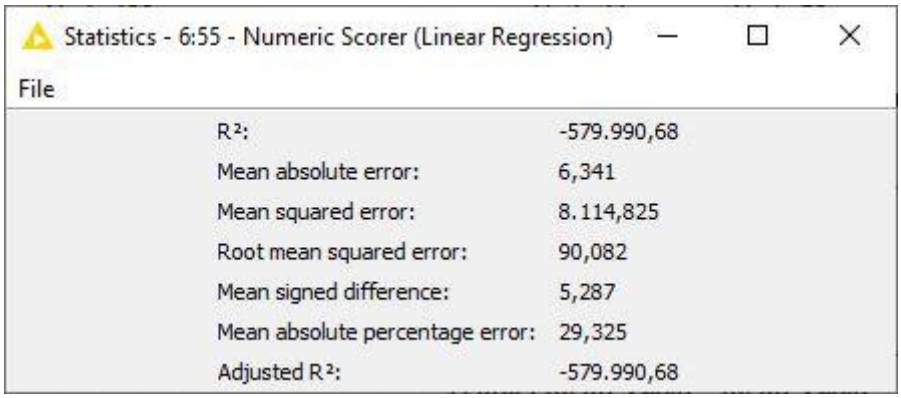


Şekil 4.58 Lineer Regresyon

Modelin öğrenmesi aşamasında Linear Regression Learner düğümüne X-Partitioner düğümünde ikiye ayrılan verinin eğitim verisi bağlanmıştır. Ayarlarda hedef değişken WS_Career olarak seçilmiştir. Tahmin düğümüne model bilgisi öğrenme düğümü olan Linear Regression Learner'den, tahmin edilecek veri bilgisi de X-Partitioner düğümünde ayrılan verisinden test kısmı gelecek şekilde bağlanmış ve çalıştırılmıştır.

Regression Predictor düğümüne performans ölçümü için kullanılan Scorer düğümü'nden önce X-Aggregator Düğümü bağlanmıştır. Böylece döngü bu kısımda bitirilmiştir. Scorer düğümünün ayarlarında gerçek değer olan "first column" "WS_Career", tahmin olan "second column" ise Prediction (WS_Career) olarak ayarlanmıştır.

Scorer sonuçlarına bakıldığında R-Kare değeri -579.990,68, MAE değeri 6,341, MSE değeri 8.114,825 RMSE değeri 90,082, MAPE değeri 29,325 olarak bulunmuştur (Şekil 4.59).



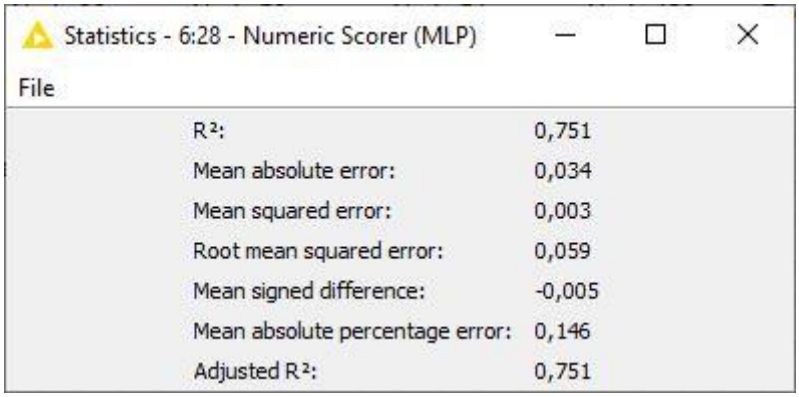
Şekil 4.59 Lineer Regresyon Sonuç

4.6.22.4. Yapay Sinir Ağları (MLP) Cross Validation

Modelin öğrenmesi aşamasında RProp MLP Learner düğümüne X-Partitioner düğümünde ikiye ayrılan verinin eğitim verisi bağlanmıştır. Ayarlarda hedef değişken WS_Career olarak seçilmiştir. Tahmin düğümüne model bilgisi öğrenme düğümü olan RProp MLP Learner'den, tahmin edilecek veri bilgisi de X-Partitioner düğümünde ayrılan verisinden test kısmı gelecek şekilde bağlanmış ve çalıştırılmıştır.

MLP Predictor düğümüne performans ölçümü için kullanılan Scorer düğümü'nden önce X-Aggregator Düğümü bağlanmıştır. Böylece döngü bu kısımda bitirilmiştir. Scorer düğümünün ayarlarında gerçek değer olan "first column" "WS_Career", tahmin olan "second column" ise Prediction (WS_Career) olarak ayarlanmıştır.

Scorer sonuçlarına bakıldığında R-Kare değeri 0,751, MAE değeri 0,034, MSE değeri 0,003 RMSE değeri 0,059, MAPE değeri 0,146 olarak bulunmuştur (Şekil 4.60).



Şekil 4.60 MLP Sonuçları

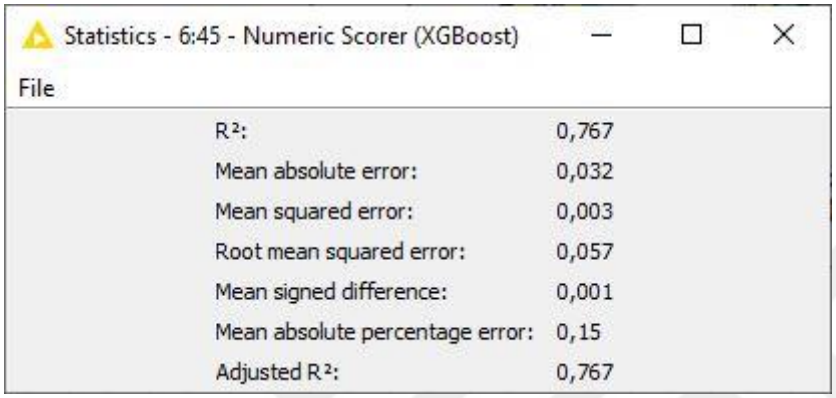
4.6.22.5. XgBoost Çapraz Doğrulama

Modelin öğrenmesi aşamasında XGBoost Tree Ensemble Learner düğümüne X-Partitioner düğümünde ikiye ayrılan verinin eğitim verisi bağlanmıştır. Ayarlarda hedef değişken WS_Career olarak seçilmiştir. Tahmin düğümüne model bilgisi öğrenme düğümü

olan XGBoost Tree Ensemble Learner'den, tahmin edilecek veri bilgisi de X-Partitioner düğümünde ayrılan verisinden test kısmı gelecek şekilde bağlanmış ve çalıştırılmıştır.

XGBoost Tree Ensemble Predictor düğümüne performans ölçümü için kullanılan Scorer düğümü'nden önce X-Aggregator Düğümü bağlanmıştır. Böylece döngü bu kısımda bitirilmiştir. Scorer düğümünün ayarlarında gerçek değer olan "first column" "WS_Career", tahmin olan "second column" ise Prediction (WS_Career) olarak ayarlanmıştır.

Scorer sonuçlarına bakıldığında R-Kare değeri 0,767, MAE değeri 0,032, MSE değeri 0,003, RMSE değeri 0,057, MAPE değeri 0,15 olarak bulunmuştur (Şekil 4.61).



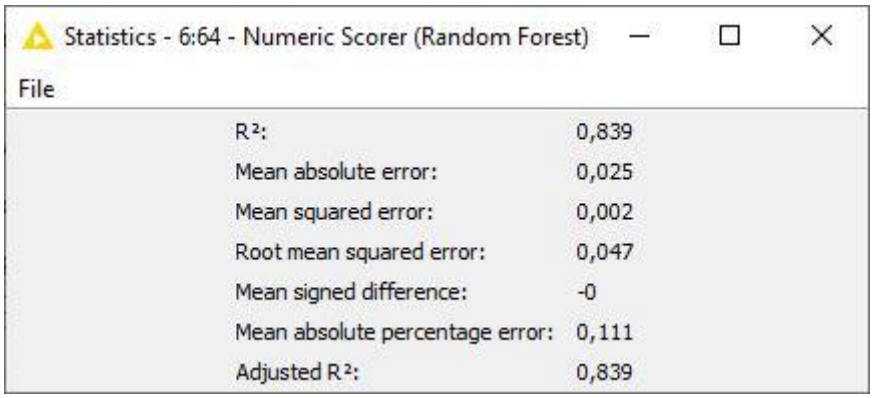
Şekil 4.61 XgBoost Learning Sonuçları

4.6.22.6. Rastgele Orman Cross Validation

Modelin öğrenmesi aşamasında Random Forest Learner düğümüne X-Partitioner düğümünde ikiye ayrılan verinin eğitim verisi bağlanmıştır. Ayarlarda hedef değişken WS_Career olarak seçilmiştir Tahmin düğümüne model bilgisi öğrenme düğümü olan Random Forest Learner'den, tahmin edilecek veri bilgisi de X-Partitioner düğümünde ayrılan verisinden test kısmı gelecek şekilde bağlanmış ve çalıştırılmıştır.

Random Forest Predictor düğümüne performans ölçümü için kullanılan Scorer düğümü'nden önce X-Aggregator Düğümü bağlanmış ve döngü bu kısımda bitirilmiştir. Scorer düğümünün ayarlarında gerçek değer olan "first column" "WS_Career", tahmin olan "second column" ise Prediction (WS_Career) olarak ayarlanmıştır.

Scorer sonuçlarına bakıldığında R-Kare değeri 0,839, MAE değeri 0,025, MSE değeri 0,002 RMSE değeri 0,047, MAPE değeri 0,111 olarak bulunmuştur (Şekil 4.62).



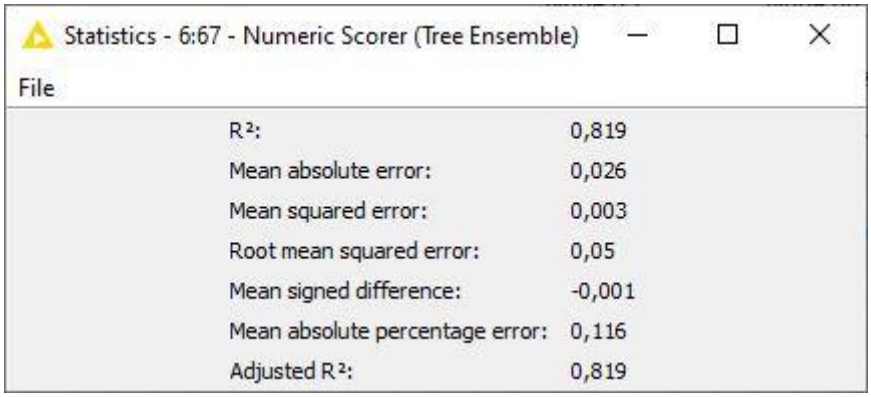
Şekil 4.62 Rastgele Orman Sonuçları

4.6.22.7. Karar Ağaçları Topluluk Öğrenmesi Rastgele Orman Çapraz Doğrulama

Modelin öğrenmesi aşamasında Tree Ensemble Learner düğümüne X-Partitioner düğümünde ikiye ayrılan verinin eğitim verisi bağlanmıştır. Ayarlarda hedef değişken WS_Career olarak seçilmiştir. Tahmin düğümüne model bilgisi öğrenme düğümü olan Tree Ensemble Learner'den, tahmin edilecek veri bilgisi de X-Partitioner düğümünde ayrılan verisinden test kısmı gelecek şekilde bağlanmış ve çalıştırılmıştır.

Tree Ensemble Predictor düğümüne performans ölçümü için kullanılan Scorer düğümü'nden önce X-Aggregator Düğümü bağlanmış ve döngü bu kısımda bitirilmiştir. Scorer düğümünün ayarlarında gerçek değer olan "first column" "WS_Career", tahmin olan "second column" ise Prediction (WS_Career) olarak ayarlanmıştır.

Scorer sonuçlarına bakıldığında R-Kare değeri 0,819, MAE değeri 0,026, MSE değeri 0,003 RMSE değeri 0,05, MAPE değeri 0,116 olarak bulunmuştur (Şekil 4.63).



Şekil 4.63 Karar Ağaçları Topluluk Öğrenmesi Rastgele Orman Sonuçları

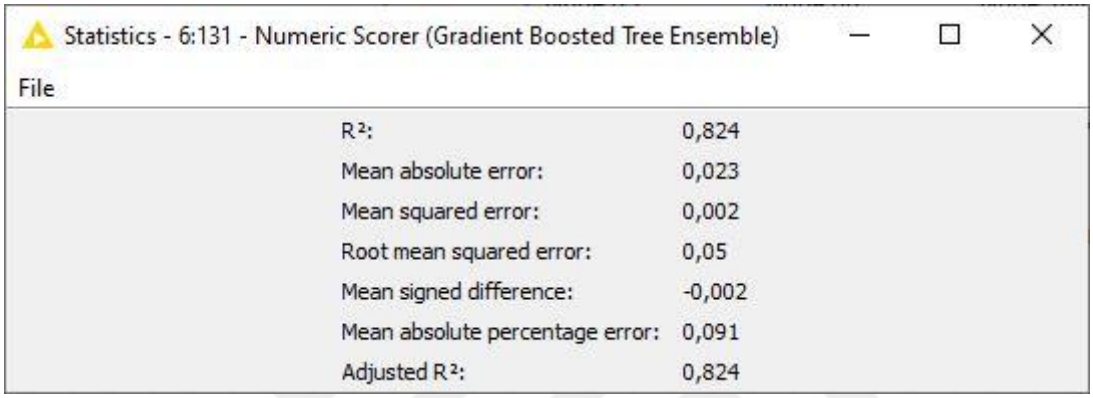
4.6.22.8. Karar Ağaçları Topluluk Öğrenmesi Gradient Boosting Çapraz Doğrulama

Modelin öğrenmesi aşamasında Gradient Boosted Trees Learner düğümüne X-Partitioner düğümünde ikiye ayrılan verinin eğitim verisi bağlanmıştır. Ayarlarda hedef değişken WS_Career olarak seçilmiştir. Tahmin düğümüne model bilgisi öğrenme düğümü

olan Gradient Boosted Trees Learner'den, tahmin edilecek veri bilgisi de X-Partitioner düğümünde ayrılan verisinden test kısmı gelecek şekilde bağlanmış ve çalıştırılmıştır.

Gradient Boosted Trees Predictor düğümüne performans ölçümü için kullanılan Scorer düğümü'nden önce X-Aggregator Düğümü bağlanmış ve döngü bu kısımda bitirilmiştir. Scorer düğümünün ayarlarında gerçek değer olan "first column" "WS_Career", tahmin olan "second column" ise Prediction (WS_Career) olarak ayarlanmıştır.

Scorer sonuçlarına bakıldığında R-Kare değeri 0,824, MAE değeri 0,023, MSE değeri 0,002 RMSE değeri 0,05, MAPE değeri 0,091 olarak bulunmuştur (Şekil 4.64).



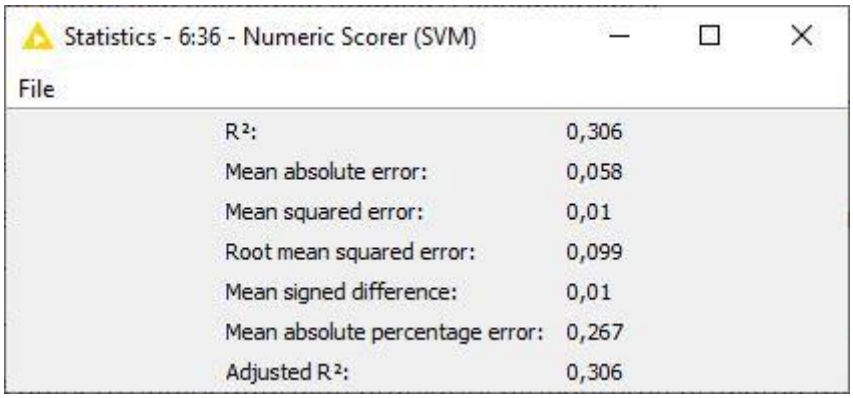
Şekil 4.64 Karar Ağaçları Topluluk Öğrenmesi Gradient Boosting Sonuçları

4.6.22.9. SVM Çapraz Doğrulama

Bu model için Number to String düğümü kullanarak Stringe çevirdiğimiz çıktı değişkeni kullanılarak modele sokulmuştur. Modelin öğrenmesi aşamasında SVM Learner düğümüne X-Partitioner düğümünde ikiye ayrılan verinin eğitim verisi bağlanmıştır. Ayarlarda hedef değişken WS_Career olarak seçilmiştir. Tahmin düğümüne model bilgisi öğrenme düğümü olan SVM Learner'den, tahmin edilecek veri bilgisi de X-Partitioner düğümünde ayrılan verisinden test kısmı gelecek şekilde bağlanmış ve çalıştırılmıştır.

SVM Predictor düğümüne performans ölçümü için kullanılan Scorer düğümü'nden önce X-Aggregator Düğümü bağlanmış ve döngü bu kısımda bitirilmiştir. Scorer düğümünün ayarlarında gerçek değer olan "first column" "WS_Career", tahmin olan "second column" ise Prediction (WS_Career) olarak ayarlanmıştır.

Scorer sonuçlarına bakıldığında R-Kare değeri 0,306, MAE değeri 0,058, MSE değeri 0,01 RMSE değeri 0,099, MAPE değeri 0,267 olarak bulunmuştur (Şekil 4.65).



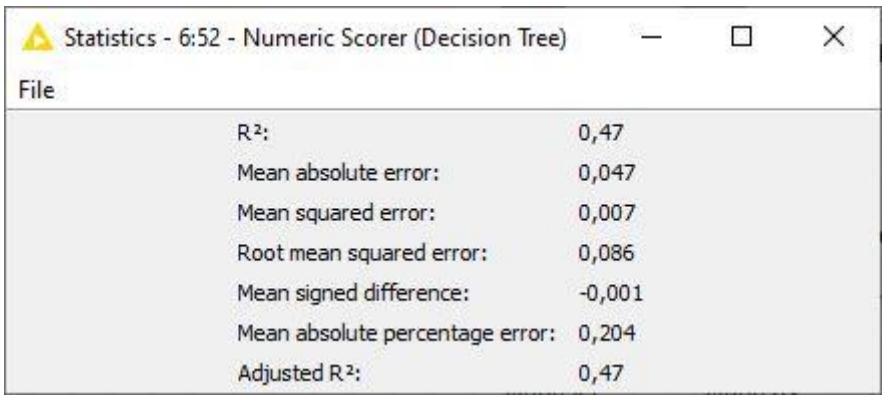
Şekil 4.65 SVM Sonuçları

4.6.22.10. Karar Ağacı Çapraz Doğrulama

Bu model için Number to String düğümü kullanarak Stringe çevirdiğimiz çıktı değişkeni kullanılarak modele sokulmuştur. Modelin öğrenmesi aşamasında Decision Tree Learner düğümüne X-Partitioner düğümünde ikiye ayrılan verinin eğitim verisi bağlanmıştır. Ayarlarda hedef değişken WS_Career olarak seçilmiştir. Tahmin düğümüne model bilgisi öğrenme düğümü olan Decision Tree Learner'den, tahmin edilecek veri bilgisi de X-Partitioner düğümünde ayrılan verisinden test kısmı gelecek şekilde bağlanmış ve çalıştırılmıştır.

Decision Tree Predictor düğümüne performans ölçümü için kullanılan Scorer düğümü'nden önce X-Aggregator Düğümü bağlanmış ve döngü bu kısımda bitirilmiştir. Scorer düğümünün ayarlarında gerçek değer olan "first column" "WS_Career", tahmin olan "second column" ise Prediction (WS_Career) olarak ayarlanmıştır.

Scorer sonuçlarına bakıldığında R-Kare değeri 0,47, MAE değeri 0,047, MSE değeri 0,007 RMSE değeri 0,086, MAPE değeri 0,204 olarak bulunmuştur (Şekil 4.66).



Şekil 4.66 Karar Ağacı Sonuçları

BEŞİNCİ BÖLÜM
BULGULAR

5.1. Bulgular

Bu çalışmanın 4.6.6. bölümünde yer alan lineer korelasyon işlemleri sonrasında veri setindeki lineer korelasyonla değişkenler arasındaki ilişkiler incelendiğinde en yüksek Salary_21-22 ile FG, FGA, FT, FTA, PTS, MP(1), 2P(1), 2PA(1), FT(1), FTA(1), TOV(1), GS(2), MP(2), 2P(2), FT(2), FTA(2), AST(2), TOV(2), PTS(2), WS_Career değişkenleri arasında 0,60'dan yüksek korelasyon vardır. Salary_21-22 ile GS(1), FG(1), FGA(1), PTS(1), FG(2), FGA(2), 2PA(2) değişkenleri arasında 0,70-0,80 arasında korelasyon vardır. Salary_21-22 ile Cap Space ve Cap Space(1) arasında negatif yönlü bir korelasyon vardır.

WS_21-22 ile Cap Space ve Cap Space(1), 3P% arasında 0,10-0,20 arasında negatif yönlü bir korelasyon vardır. WS_21-22 ile GS, FG%, 2P, DRB, TRB, 2P(2), ORB(2), DRB(2), TRB(2) değişkenleri 0,50-0,60 arasında korelasyon vardır. WS_21-22 değişkeni ile %60'dan yüksek korelasyonu olan bir değişken yoktur.

WS_Career değişkeni ile Age, FT(1), FTA(1), FG(2), FGA(2), 2P(2), 2PA(2), FT(2), FTA(2), DRB(2), AST(2), TOV(2),Twitter arasında 0,60-0,70 arasında korelasyon vardır. WS_Career değişkeni ile Experience değişkeni ile 0,70-0,80; G(1), GS(1), G(2), GS(2) arasında 0,80-0,90 arasında korelasyon vardır.

WS_Career değişkeniyle tüm salary değişkenleri arasında 0,60-0,70 arasında korelasyon vardır.

Veri setine eklenen bütün değişkenler için bu çalışmanın 4.6.2 bölümünde yer alan histogramları incelendiğinde win share değişkenleri normal dağılım göstermiştir.

Bu bulgular ışığında WS yani win share değişkenleriyle Salary'ler açıklanmaya çalışılmıştır. Bu bölümde WS_Career çıktısına uygulanan modellerin değerlendirilmesi yapılacaktır. Sırasıyla Lineer Regresyon, Yapay Sinir Ağları, XgBoost Learning, Rastgele Orman, Karar Ağaçları Topluluk Öğrenmesi, SVM, Karar Ağaçları için üstte eğitim altta test verilerinin performansları bulunmaktadır.

R Kare bağımsız değişkenlerin bağımlı(hedef) değişkendeki varyansı açıklama yeteneğidir ve 1'e ne kadar yakınsa, bağımsız değişkenlerin bağımlı(hedef) değişkendeki varyansı açıklama yeteneği o kadar yüksek ve model hassas demektir olarak yorumlanacaktır. Düzeltilmiş R^2, birden fazla bağımsız değişkenin kullanıldığı regresyon analizlerinde R^2'nin etkileşimleri hesaba katması için düzenlenmiş bir istatistik değeridir (Barut ve Bilgin, 2023: 155). Düzeltilmiş R^2 ise birden fazla bağımsız değişkenin eklenmesiyle R^2'nin yükselebileceği

gerçeğini de dikkate alır ve bu nedenle modelin gerçekçi bir değerlendirmesi için daha uygun bir ölçüttür ancak modeldeki tüm modellerde R-Kare ve D Düzeltilmiş R-Kare değerleri aynı çıktığı için değerlendirmede sadece R-Kare değerine bakılacaktır.

Ortalama Mutlak Hata (Mean Absolute Error, MAE) bir regresyon modelinin tahminlerinin gerçek değerlerden ne miktarda uzak olduğunu ölçmek için kullanılan bir metriktir (Özen vd., 2021: 136). Daha düşük bir MAE değeri modelin daha iyi tahminler yaptığını ve gerçek değerlere daha yakın olduğunu gösterir.

Ortalama Kare Hatası (Mean Squared Error, MSE), MAE gibi bir regresyon modelinin tahminlerinin gerçek değerlerden uzaklığını ölçmek için kullanılmaktadır. MSE'nin orijinal veri birimleri üzerinde kareler alınarak hesaplanması nedeniyle yorumu bazen zor olabilmektedir. Bu nedenle RMSE (Kök Ortalama Kare Hata) gibi daha yorumlanabilir alternatifler de tercih edilmektedir (Hodson, 2022: 5481-5487).

Kök Ortalama Kare Hatası (Root Mean Squared Error, RMSE), MSE'nin karekökünü alarak elde edilen bir regresyon performans metriğidir (Özen vd., 2021: 137). RMSE değerinin genellikle daha küçük olması istenmektedir. Daha düşük MSE ve RMSE değeri modelin daha iyi tahminler yaptığını ve gerçek değerlere daha yakın olduğunu göstermektedir.

Lineer Regresyon Eğitim sonuçlarına bakıldığında R-Kare değeri 0,981, MAE değeri 0,012, MSE değeri 0, RMSE değeri 0,016, MAPE değeri 0,071 olarak bulunmuştur.

Lineer Regresyon Test sonuçlarına bakıldığında R-Kare değeri-1.200.783,81, MAE değeri 19,157, MSE değeri 17,574,184, RMSE değeri 132,568, MAPE değeri 86,349 olarak bulunmuştur.

Yapay Sinir Ağları Eğitim sonuçlarına bakıldığında R-Kare değeri 0,999, MAE değeri 0,003, MSE değeri 0, RMSE değeri 0,004, MAPE değeri 0,019 olarak bulunmuştur.

Yapay Sinir Ağları Test sonuçlarına bakıldığında R-Kare değeri 0,543, MAE değeri 0,038, MSE değeri 0,007, RMSE değeri 0,082, MAPE değeri 0,151 olarak bulunmuştur.

XgBoost Learning Eğitim sonuçlarına bakıldığında R-Kare değeri 0,851, MAE değeri 0,017, MSE değeri 0,002, RMSE değeri 0,045, MAPE değeri 0,084 olarak bulunmuştur.

XgBoost Learning Test sonuçlarına bakıldığında R-Kare değeri 0,868, MAE değeri 0,028, MSE değeri 0,002, RMSE değeri 0,044, MAPE değeri 0,129 olarak bulunmuştur.

Rastgele Orman Eğitim sonuçlarına bakıldığında R-Kare değeri 0,973, MAE değeri 0,01, MSE değeri 0, RMSE değeri 0,019, MAPE değeri 0,045 olarak bulunmuştur.

Rastgele Orman Test sonuçlarına bakıldığında R-Kare değeri 0,91, MAE değeri 0,024, MSE değeri 0,001, RMSE değeri 0,036, MAPE değeri 0,098 olarak bulunmuştur.

Karar Ağaçları Topluluk Öğrenmesi Rastgele Orman Eğitim sonuçlarına bakıldığında R-Kare değeri 0,968, MAE değeri 0,011, MSE değeri 0, RMSE değeri 0,021, MAPE değeri 0,046 olarak bulunmuştur.

Karar Ağaçları Topluluk Öğrenmesi Rastgele Orman Test sonuçlarına bakıldığında R-Kare değeri 0,9, MAE değeri 0,022, MSE değeri 0,001, RMSE değeri 0,038, MAPE değeri 0,09 olarak bulunmuştur.

Karar Ağaçları Topluluk Öğrenmesi Gradient Boosting Eğitim sonuçlarına bakıldığında R-Kare değeri 1, MAE değeri 0,001, MSE değeri 0, RMSE değeri 0,002, MAPE değeri 0,003 olarak bulunmuştur.

Karar Ağaçları Topluluk Öğrenmesi Gradient Boosting Test sonuçlarına bakıldığında R-Kare değeri 0,88, MAE değeri 0,022, MSE değeri 0,002, RMSE değeri 0,042, MAPE değeri 0,089 olarak bulunmuştur.

SVM Eğitim sonuçlarına bakıldığında R-Kare değeri 0,614, MAE değeri 0,059, MSE değeri 0,09, RMSE değeri 0,093, MAPE değeri 0,243 olarak bulunmuştur.

SVM Test sonuçlarına bakıldığında R-Kare değeri 0,999, MAE değeri 0, MSE değeri 0, RMSE değeri 0,003, MAPE değeri 0,002 olarak bulunmuştur.

Karar Ağaçları Eğitim sonuçlarına bakıldığında R-Kare değeri 0,609, MAE değeri 0,054, MSE değeri 0,009, RMSE değeri 0,094, MAPE değeri 0,209 olarak bulunmuştur.

Karar Ağaçları Test sonuçlarına bakıldığında R-Kare değeri 0,75, MAE değeri 0,023, MSE değeri 0,03, RMSE değeri 0,054, MAPE değeri 0,101 olarak bulunmuştur.

Tablo 5.1 Modellerin Değerlendirilmesi

Model	R-Kare	MAE	MSE	RMSE	MAPE
Lineer Regresyon Eğitim	0,981	0,012	0	0,016	0,071
Lineer Regresyon Test	-1200783,81	19,157	17574,184	132,568	86,349
Yapay Sinir Ağları Eğitim	0,999	0,003	0	0,004	0,019
Yapay Sinir Ağları Test	0,543	0,038	0,007	0,082	0,151
XgBoost Learning Eğitim	0,851	0,017	0,002	0,045	0,084
XgBoost Learning Test	0,868	0,028	0,002	0,044	0,129
Rastgele Orman Eğitim	0,973	0,01	0	0,019	0,045
Rastgele Orman Test	0,91	0,024	0,001	0,036	0,098
Karar Ağaçları Topluluk Öğrenmesi Rastgele Orman Eğitim	0,968	0,011	0	0,021	0,046
Karar Ağaçları Topluluk Öğrenmesi Rastgele Orman Test	0,9	0,022	0,001	0,038	0,09

Karar Ağaçları Topluluk Öğrenmesi Gradient Boosting Eğitim	1	0,001	0	0,002	0,003
Karar Ağaçları Topluluk Öğrenmesi Gradient Boosting Test	0,88	0,022	0,002	0,042	0,089
SVM Eğitim	0,614	0,059	0,09	0,093	0,243
SVM Test	0,999	0	0	0,003	0,002
Karar Ağaçları Eğitim	0,609	0,054	0,009	0,094	0,209
Karar Ağaçları Test	0,75	0,023	0,03	0,054	0,101

Bu tabloya göre test verilerine bakıldığında model başarısı en yüksek modeller R Kare'ye göre 0,91 ile Rastgele Orman, 0,9 ile Karar Ağaçları Topluluk Öğrenmesi Rastgele Orman, 0,868 ile XgBoost Learning olarak tespit edilmiştir. SVM modelinin test aşamasında, Karar Ağaçları Topluluk Öğrenmesi Gradient Boosting eğitim aşamasında ezberleme yaptığı düşünüldüğü için bu sıralamaya dahil edilmemiştir.

Yüzde 85 başarı oranını yakalamak istendiği için bu veri seti için en uygun Rastgele Orman ve Karar Ağaçları Topluluk Öğrenmesi Rastgele Orman gibi gözükse de eğitim sonuçlarında MSE değerleri 0'dır. Aynı zamanda bu modellerde test ve eğitim verileri arasındaki sapma XgBoost Learning modelinden çok daha fazladır. XgBoost Learning bu veri için en uygun modeldir.

Model Doğrulama Sonrası Performanslar

Model doğrulama için yapılan Çapraz Doğrulama sonrası modellerin performansları incelenmiştir.

Lineer Regresyon sonuçlarına bakıldığında R-Kare değeri -579.990,68, MAE değeri 6,341, MSE değeri 8.114,825 RMSE değeri 90,082, MAPE değeri 29,325 olarak bulunmuştur. Sonuçlar öncekine göre iyileşmiş olsa da Lineer Regresyon bu modele uygun değildir.

Yapay Sinir Ağları sonuçlarına bakıldığında R-Kare değeri 0,751, MAE değeri 0,034, MSE değeri 0,003 RMSE değeri 0,059, MAPE değeri 0,146 olarak bulunmuştur. Çapraz Doğrulama yapılan veride yapay sinir ağlarında performans artmıştır.

XgBoost sonuçlarına bakıldığında R-Kare değeri 0,767, MAE değeri 0,032, MSE değeri 0,003, RMSE değeri 0,057, MAPE değeri 0,15 olarak bulunmuştur. Çapraz Doğrulama yapılan veride XgBoost modelinde performans düşmüştür.

Rastgele Orman sonuçlarına bakıldığında R-Kare değeri 0,839, MAE değeri 0,025, MSE değeri 0,002 RMSE değeri 0,047, MAPE değeri 0,111 olarak bulunmuştur. Çapraz Doğrulama yapılan veride Rastgele orman modelinde performans minimal düzeyde düşmüştür.

Karar Ağaçları Topluluk Öğrenmesi Rastgele Orman sonuçlarına bakıldığında R-Kare değeri 0,819, MAE değeri 0,026, MSE değeri 0,003 RMSE değeri 0,05, MAPE değeri 0,116

olarak bulunmuştur. Çapraz Doğrulama yapılan veride Karar Ağaçları Topluluk Öğrenmesi Rastgele Orman modelinde performans minimal düzeyde düşmüştür.

Karar Ağaçları Topluluk Öğrenmesi Gradient Boosting sonuçlarına bakıldığında R-Kare değeri 0,824, MAE değeri 0,023, MSE değeri 0,002 RMSE değeri 0,05, MAPE değeri 0,091 olarak bulunmuştur. Çapraz Doğrulama yapılan veride Karar Ağaçları Topluluk Öğrenmesi Gradient Boosting modelinde performans minimal düzeyde düşmüştür.

SVM sonuçlarına bakıldığında R-Kare değeri 0,306, MAE değeri 0,058, MSE değeri 0,01 RMSE değeri 0,099, MAPE değeri 0,267 olarak bulunmuştur. Çapraz Doğrulama yapılan veride SVM modelinde performans düşmüş görünüyor fakat önceki modelde ezberleme yaptığı düşünülen SVM aşırı öğrenmeden kurtulmuştur.

Karar Ağaçları sonuçlarına bakıldığında R-Kare değeri 0,47, MAE değeri 0,047, MSE değeri 0,007 RMSE değeri 0,086, MAPE değeri 0,204 olarak bulunmuştur. Çapraz Doğrulama yapılan veride Karar Ağaçları modelinde performans düşmüştür.

Tablo 5. 2 İyileştirme Sonrası Modellerin Değerlendirilmesi

Model	R-Kare	MAE	MSE	RMSE	MAPE
Lineer Regresyon Eğitim	0,981	0,012	0	0,016	0,071
Lineer Regresyon Test	-1200783,81	19,157	17574,184	132,568	86,349
Lineer Regresyon-Çapraz Doğrulama	**-579990,68**	**6,341**	**8.114**	**90,082**	**29,325**
Yapay Sinir Ağları Eğitim	0,999	0,003	0	0,004	0,019
Yapay Sinir Ağları Test	0,543	0,038	0,007	0,082	0,151
Yapay Sinir Ağları- Çapraz Doğrulama	**0,751**	**0,034**	**0,003**	**0,059**	**0,146**
XgBoost Learning Eğitim	0,851	0,017	0,002	0,045	0,084
XgBoost Learning Test	0,868	0,028	0,002	0,044	0,129
XgBoost- Çapraz Doğrulama	**0,767**	**0,032**	**0,003**	**0,057**	**0,15**
Rastgele Orman Eğitim	0,973	0,01	0	0,019	0,045
Rastgele Orman Test	0,91	0,024	0,001	0,036	0,098
Rastgele Orman- Çapraz Doğrulama	**0,839**	**0,025**	**0,002**	**0,047**	**0,111**
Karar Ağaçları Topluluk Öğrenmesi Rastgele Orman Eğitim	0,968	0,011	0	0,021	0,046
Karar Ağaçları Topluluk Öğrenmesi Rastgele Orman Test	0,9	0,022	0,001	0,038	0,09
Karar Ağaçları Topluluk Öğrenmesi Rastgele Orman- Çapraz Doğrulama	**0,819**	**0,026**	**0,003**	**0,05**	**0,116**
Karar Ağaçları Topluluk Öğrenmesi Gradient Boosting Eğitim	1	0,001	0	0,002	0,003
Karar Ağaçları Topluluk Öğrenmesi Gradient Boosting Test	0,88	0,022	0,002	0,042	0,089

Karar Ağaçları Topluluk Öğrenmesi Gradient Boosting- Çapraz Doğrulama	**0,824**	**0,023**	**0,002**	**0,05**	**0,091**
SVM Eğitim	0,614	0,059	0,09	0,093	0,243
SVM Test	0,999	0	0	0,003	0,002
SVM- Çapraz Doğrulama	**0,306**	**0,058**	**0,01**	**0,099**	**0,267**
Karar Ağaçları Eğitim	0,609	0,054	0,009	0,094	0,209
Karar Ağaçları Test	0,75	0,023	0,03	0,054	0,101
Karar Ağaçları- Çapraz Doğrulama	**0,47**	**0,047**	**0,007**	**0,086**	**0,204**

Bu tabloya göre doğrulama sonrası model başarısı en yüksek modeller R Kare'ye göre **0,839** ile Rastgele Orman, **0,824** ile Karar Ağaçları Topluluk Öğrenmesi Gradient Boosting, **0,819** ile Karar Ağaçları Topluluk Öğrenmesi Rastgele Orman olarak tespit edilmiştir.

Literatür taraması sonrasında yüzde 85 başarı oranı yakalamak istenildiğinden dolayı bu veri setine en uygun Rastgele Orman, Karar Ağaçları Topluluk Öğrenmesi Gradient Boosting ve Karar Ağaçları Topluluk Öğrenmesi Rastgele Orman bu veri için en uygun modeldir.

SONUÇ

Geçmiş verilere bakılarak gelecekte olabilecek olayları tahmin etmek birçok kurum ve bilim alanı için önem arz etmektedir. Alanda istatiksel olarak NBA oyuncuları ve oyunlarla ilgili çeşitli tahminlemelerde bulunulan çalışmalar yer almaktadır. Makine öğrenmesi yöntemleri kullanılarak NBA oyuncularının istatistiki verilerinden tahminleme yapan çalışmalar ise literatürde son yıllarda yer almaya başlamıştır. Bu çalışmaları devam ettirmek amacıyla makine öğrenmesi yöntemlerinin birçoğunu kullanarak makine öğrenmesi modellerini o veriler üzerinde performanslarını ölçümleme yapmak ve bulguları analiz ederek makine öğrenmesi modellerinin başarıları ve Win Share değişkeninin maaş hesaplamadaki önemi literatüre kazandırmak istenilmiştir. Çıktı değişkeni olan Win Share terimi sporcuların veya takımların bir yarışmada veya oyunun sonucunda kazandığı veya kazanmayı başardığı başarı payını ifade eder (Basketball Reference, 2023b).

Win Share formülde bu değişkenler içerdiği ve maaşı da yüzde 60 korelasyonla etkilediği için bu çalışmada kullanılarak Win Share hesaplanmıştır. Win Share değişkenini sekiz farklı makine öğrenmesi algoritması kullanarak tahmin ettirilmiştir. Bu algoritmalar içinden Win Share'i en başarılı şekilde tahmin eden algoritmalar tespit edilmiştir. En iyi başarılı tahminleme yapan algoritma kullanılarak bir maaş hesabına ulaşılabilecektir. Bu çalışmada hesaplanan Win Share kullanılarak gerçek bir değer elde edilebilecektir. Bu şekilde oyuncuların hak ettiği maaşa yakın gerçek bir bulgu elde edilebilir. Bu bulgu kullanılarak spor kulüplerine yardım edilebilir ayrıca sektör gerçek değerleri mi veriyor kıyaslama imkânı doğacaktır. Bu oyuncuların kelepir mi yoksa yüksek ödeme mi aldıklarına dair daha net sonuçlara ulaşılabilmesi için tahmin edilmeye çalışılmıştır.

Tez çalışmasında 2021-2022 sezonu NBA oyuncularının çeşitli istatistiklerinin WS_Career yani oyuncuların kariyerleri boyunca takımlara kazandırma oranlarına olan etkisinin makine öğrenmesi yöntemleriyle analizi yapılmıştır. Veri seti WS_Career çıktı değişkeni olmak üzere toplamda 98 değişken ve 240 gözlemden oluşturulmuştur. Tezin analiz sürecinde Knime programı kullanılmıştır. Veri setinde ön işleme adımında 9 değişken veri setinden çıkartılmıştır. Normalizasyona ihtiyaç duyan algoritmalar kullanıldığı için veri setinde tüm değişkenlere normalizasyon gerçekleştirilmiştir. Lineer Regresyon, Yapay Sinir Ağları, XgBoost Learning, Rastgele Orman, Karar Ağaçları Topluluk Öğrenmesi Rastgele Orman, Karar Ağaçları Topluluk Öğrenmesi Gradient Boosting, SVM, Karar Ağaçları olmak üzere toplam sekiz algoritma kullanılmıştır. Makine öğrenmesi algoritmalarında performanslar direk

olarak algoritmaların seçilen parametrelerine bağlı olduğu için doğru parametrenin seçilmesine yönelik çalışma gerçekleştirilmiştir. En iyi performansı veren parametreler üzerinden algoritmalar çalıştırılmıştır. Değerlendirme ölçütü olarak bu modellerin R-Kare, MAE, MSE, RMSE ve MAPE hata ölçüm sonuçları karşılaştırılarak tablolaştırılmıştır. XgBoost Learning algoritmasının bu veri için en uygun model olduğu tespit edilmiştir. Tüm algoritmalar için 5 Kat Çapraz Doğrulama tekniği kullanılmıştır. En iyi performansı veren parametreler ile bu modeller çalıştırılmıştır. Çapraz Doğrulama sonrası bu veri seti için en uygun Rastgele Orman, Karar Ağaçları Topluluk Öğrenmesi Gradient Boosting ve Karar Ağaçları Topluluk Öğrenmesi Rastgele Orman modelleri olarak tespit edilmiştir.

Oyuncuların sezon verileri üzerine uygulanan lineer korelasyon sonucunda maaş ve Win Share değişkenleri arasında yüzde 60 üzerinde bir korelasyon olduğu belirlenmiştir. Bu iki değişken arasındaki ilişki ile Win Share'i açıklamanın maaş değişkenini belirlemede önemli bir ana faktör olacağı düşünülmektedir.

WS_Career değişkeni ile Yaş, Kariyerinde Serbest Atış Sayısı, Kariyerinde Serbest Atış Deneme Sayısı, Playoff Saha İçi İsabet, Playoff Saha İçi Denemeleri (Ortalama), Playoff İki Sayılık Atış Sayısı, Playoff İki Sayılık Atış Deneme Sayısı, Playoff Serbest Atış Sayısı, Playoff Serbest Atış Deneme Sayısı, Playoff Defansif/Savunma Ribaunt, Playoff Asist, Playoff Top Kaybı, Twitter arasında yüzde 60 ve 70 arasında korelasyon vardır. WS_Career değişkeni ile Deneyim değişkeni ile yüzde 70 ve 80 aralığında; Kariyerinde Oynadığı Maç Sayısı, Kariyerinde Başlangıç Oyuncusu Olunan Maç Sayısı, Playoff Oynadığı Maç Sayısı, Playoff Başlangıç Oyuncusu Olunan Maç Sayısı arasında yüzde 80 ve 90 arasında korelasyon vardır. Tüm bu değişkenler Win Share'i etkilemektedir. Win Share değişkenine etkisi olacağı düşünülerek veri setine eklenen oyuncuların ve takımın sosyal medya hesapları olan Twitter ve Instagram'dan sadece Twitter'ın yüzde 60 üzerinde bir korelasyonla tek yönlü olarak Win Share değişkenini etkilediği belirlenmiştir.

Algoritmaların karşılaştırmalı sonuç tablosuna göre doğrulama sonrası model başarısı en yüksek modeller R Kare'ye göre 0,839 ile Rastgele Orman, 0,824 ile Karar Ağaçları Topluluk Öğrenmesi Gradient Boosting, 0,819 ile Karar Ağaçları Topluluk Öğrenmesi Rastgele Orman olarak tespit edilmiştir. Rastgele Orman'ın MAPE değeri olan yüzde 11,4 mutlak tahmin sapması ve Karar Ağaçları Topluluk Öğrenmesi Rastgele Orman algoritması yüzde 11,6 tahmin sapmasıyla doğru tahmin modelleri oldukları tespit edilmiştir. Karar Ağaçları Topluluk Öğrenmesi Gradient Boosting algoritması yüzde 9 tahmin sapmasıyla yüksek doğruluk derecesine sahip bir model olmuştur.

Kullanılan algoritmalar ile tahmin edilen Win Share değişkenini ana değişken olarak kullanarak oyuncu maaş hesaplaması yapılabilecektir. Tahminlenen Win Share üzerinden makine öğrenmesi yöntemleri kullanarak elde edilen maaş oyuncunun gerçekten hak ettiği maaş tahmini olacaktır. Bu yöntem izlenerek oyuncuların aldığı maaşların hak ettiği gerçek maaşın altında mı yoksa hak ettiğinin üzerinde mi olduğu tespit edilebilecektir. Hak ettiğinden fazla alan ve altında kalan oyuncular belirlenebilecektir.

Çalışmanın 1.2.1. bölümünde bahsedildiği gibi NBA'in takımlara uyguladığı maaş tavanı olmasına rağmen bazı durumlarda bu tavan aşılabilmektedir. Takımların bütçesi büyük takımlarda daha yüksek rakamlara ulaşabilmektedir (Ötkan ve Çolakoğlu, 2017). Bu takımlarda yer alan oyuncuların da daha yüksek maaşlar alabilmesi mümkün olmaktadır. Ancak James Harden gibi örneklerde oyuncular her zaman yüksek maaşı tercih etmemekte hak ettikleri değeri koruyarak farklı motivasyonlarla büyük takımlardan daha düşük maaşlara sözleşme imzalamaktadırlar. Bu noktada takımların bu oyuncuları elinde tutabilmesi ve oyuncuların da değerini bilerek hareket edebilmesi için oyuncuların takımlara kazandırma oranları üzerinden hak ettikleri maaşları bilmek önem arz etmektedir.

Oyuncunun hak ettiği maaşı mı istediği, takımı zor durumda bırakmaya çalışıp çalışmaması, istenilen maaşın marka değerinden mi kaynaklandığı gibi sorulara yanıt olabilmesi açısından Win Share'i doğru hesaplamak önemlidir. Örneğin NBA oyuncusu James Harden 9 yıl oynadığı Houston Rockets takımından 2020-2021 sezonunun ilk aylarında kendisi gibi süper yıldız oyuncu olan Kevin Durant ve Kyrie Irving Brooklyn'in oynadığı Nets takımına takasını istemiştir. Nets takımı Durant, Kyrie ve Harden'dan oluşan yıldız oyuncularla bir büyük üçlü oluşturmuştur ancak takım beklenen yüksek performansı göstermemiştir. James Harden 2021-22 sezonunda tekrar takasını istemiştir. Harden takımdan ayrılmak istemesine sebep olarak Nets takımında oturmuş bir sistemin olmamasını ve en iyi oyuncu olmak için bu yapının gerekli olduğunu söylemiştir. Sonraki sezonda Brooklyn Nets takımındaki diğer süper yıldız oyuncu Kevin Durant da takımdan ayrılmak istemiştir. Harden Brooklyn Nets'ten Sixers takımına gittikten sonra maaşında 14 milyon dolarlık indirime giderek yeni bir sözleşme imzalamıştır. Oyuncu maaşında indirim yapmasına rağmen hakkının verilmediğini söylemiştir (Eurohoops.net, 2022). Oyuncuların hak ettikleri maaşları net olarak bilinmesinin karar verme süreçlerinde takım ve oyuncular için etkili olacaktır. Harden Houston Rockets takımında oynadığı 9 yılda hiç şampiyonluk kazanmamıştır. Şampiyon olabilmek için takas isteyen Harden'ın iyi oyuncularla birlikte oynayabilmek amacıyla yıllardır oynadığı takımdan ayrılmak isteyerek Nets takımına gelmesi oyuncuların maaş dışındaki motivasyonuna örnek

gösterilebilir. Harden Brooklyn Nets takımında şampiyon olamayınca maaşını azaltıp büyük takımdan ayrılmak isteyerek kazandıran takıma gitmeyi tercih etmiştir.

Başka bir NBA oyuncusu Shaquille O'Neal oynadığı Orlando Magic takımıyla sözleşme yenilemeyerek şampiyon olmak için Los Angeles Lakers takımıyla sözleşme imzalamıştır. Orlando Magic takımı lüks vergisini de ödemeyi planlayarak Lakers takımının vereceği maaşın benzer miktarını teklif etmesine rağmen O'Neal daha büyük bir şehre gitmeyi ve şampiyon olabilecek bir takımda oynamak istediğini söylemiştir (Scope Medya A.Ş, 2016). İyi oyuncuların büyük takımların bütçelerinden iyi maaşlar elde etmek yerine şampiyonluk kazanmak ya da iyi oyuncularla birlikte oynamak için şampiyonluk kazandıracaklarını düşündükleri takımlara gidebileceğinin örneklerine rastlanmaktadır.

Yapılan literatür taramasında maaş hesaplama ile ilgili çalışmalarda Win Share değerine rastlanmamıştır. Win Share'in maaş üzerinde etkisinin daha önce incelenmemiş olduğu tespit edilmiştir. Bu tez çalışmasıyla Win Share' in maaş hesaplamada kullanılabilecek önemli bir değişken olmasıyla literatüre katkıda bulunacağı düşünülmektedir.

Oyuncuların istatistiklerinin Win Share'e etkisinin birden çok değişkenle çok sayıda makine öğrenmesi teknikleriyle analiz edilmesi literatürde yer alan çalışmalardan farklılık göstermektedir. Hangi modellerin ve girdi setlerinin en uygun olduğunu anlamak için farklı makine öğrenmesi algoritmaları karşılaştırılmış ve farklı girdi değişkenleri kullanılmıştır. En iyi maaş tahmini çalışmalarının çoğu yalnızca belirli bir modele odaklanır, dolayısıyla bu Win Share tahmini modellemesi için hangi modellerin ve girdilerin önemli olduğunu göstermek açısından faydalı olabilir. Kullanılan her bir algoritma için en yüksek doğruluk oranına ulaşılan parametre optimizasyonları çalışmada gösterilmiştir. Değerlendirme ölçütü olarak bu modellerin tüm R-Kare, MAE, MSE, RMSE ve MAPE sonuçlarının tamamına yer verilmiştir. Modellerdeki en yüksek doğruluğu sağlayan parametre optimizasyonlarının ve en yüksek başarı sağlayan modellerin alanda katkı sağlayabileceği düşünülmektedir.

Gelecek çalışmayla oyuncuların Win Share üzerinden maaş hesaplanmasıyla şirketlerin oyuncunun hak ettiği gerçek maaşı bilmesi önemli bir bilgi olacaktır. Oyuncunun hak ettiği maaşı bilen takım ve şirketler oyuncuların istediği maaşı verme kararında daha gerçekçi sonuçlara ulaşabilecektir. Bu makine öğrenmesi teknikleri ile spor endüstrisindeki uygulamalar daha da geliştirilebilir.

Bu çalışmanın sonuçları NBA ve diğer spor ligleri için maaş yönetimi stratejilerini iyileştirme ve oyuncu değerlendirmelerini daha bilinçli bir şekilde yapma konusunda değerli bilgiler sağlayabilir. Elde edilen sonuçların, maaş tahminlerinde ve oyuncu

değerlendirmelerinde kullanılmasıyla, spor sektörüne ve diğer ilgili sektörlere büyük faydalar sağlayabileceği düşünülmektedir.

Özet olarak bu çalışmada makine öğrenmesi modelleri kullanarak elde edilen en başarılı çalışan modelin hesapladığı Win Share değişkeni ile hesaplanacak maaş oyuncunun gerçekten hak ettiği maaş tahmini olacaktır. Oyuncunun aldığı maaş ve hak ettiği maaş arasındaki farkın oyuncunun marka değeri gibi değişkenler olabileceği düşünülmektedir. Aradaki farkı etkileyen faktörlerle ilgili çalışmalar yapılabilir. Bu bulgunun literatüre katkısı olacağı düşünülmektedir.

Makine öğrenmesi algoritmalarının hepsi her veri setinde ve durumda en başarılı tahminlemelere uygun olmayabilmektedir. Spor endüstrisi metrikleri tahmin etmesi zor alanlardır. Veri setine uygun olan modeli belirlemek bu noktada alana katkı için değer kazanmaktadır. Çalışmada sekiz algoritma içinden veriye uygun en iyi tahminleme yapan algoritma bulunmuştur. Bu modeli kullanarak maaş hesaplanabilecek ve gerçek bir değer elde edilebilecektir.

Spor endüstrisinde gün geçtikçe analitik yöntemlerin önemi artmaktadır. Yapılan makine öğrenmesi çalışmalarıyla daha gerçek bulgulara ulaşılabilecektir. Oyuncuların o pozisyonda hak ettikleri parayı alıp almaması, oyuncuların kulüpleri kandırıp kandırmadığı, oyuncuların muadili olmadığı için mi kulüpleri zora sokma durumunda bıraktıkları, oyuncuların istedikleri maaşların marka değeri mi yoksa hak ettikleri maaşı mı istedikleri, kulüplerin oyuncuyu değerinden aşağıda mı ya da yüksek ücretlere mi aldığı gibi sorulara yanıt olabileceği düşünülmektedir. Çalışma oyuncuların maaşlarını belirleme sürecindeki adaleti, sözleşme düzenlemelerini değerlendirmede ve spor endüstrisindeki maaş eşitsizliklerini araştırmak için de önemli bir kaynak olabilir.

KAYNAKÇA

Abioye, E. A., Hensel, O., Esau, T. J., Elijah, O., Abidin, M. S. Z., Ayobami, A. S., . . . Nasirahmadi, A. (2022). Precision irrigation management using machine learning and digital farming solutions. *AgriEngineering, c. 4*(1), s. 70-103.

Ahmet, E. (2021). Yapay Zekâ Odaklı Siber Risk ve Güvenlik Yönetimi. *Uluslararası Yönetim Bilişim Sistemleri ve Bilgisayar Bilimleri Dergisi, c. 5*(2), s. 144-165.

Akay, E. Ç. (2018). Ekonometride Yeni Bir Ufuk: Büyük Veri ve Makine Öğrenmesi. *Sosyal Bilimler Araştırma Dergisi, c. 7*(2), s. 41-53.

Akın, Y. K. (2008). *Veri Madenciliğinde Kümeleme Algoritmaları ve Kümeleme Analizi.* Yayımlanmamış Doktora Tezi Yayımlanmamış Doktora Tezi. Marmara Universitesi Sosyal Bilimler Enstitüsü, İstanbul.

Akşehir, Z. D. ve Kılıç, E. (2019). Makine Öğrenmesi Teknikleri İle Banka Hisse Senetlerinin Fiyat Tahmini. *Türkiye Bilişim Vakfı Bilgisayar Bilimleri ve Mühendisliği Dergisi, c. 12*(2), s. 30-39.

Albayrak, A. S. ve Yilmaz, S. K. (2009). Veri Madenciliği: Karar Ağacı Algoritmaları ve İMKB Verileri Üzerine Bir Uygulama. *Suleyman Demirel University Journal of Faculty of Economics & Administrative Sciences, c. 14*(1).

Albert, A. A., de Mingo López, L. F., Allbright, K. ve Gómez Blas, N. (2021). A Hybrid Machine Learning Model for Predicting USA NBA All-Stars. *Electronics, c. 11*(1), s. 97.

Ali, J., Khan, R., Ahmad, N. ve Maqsood, I. (2012). Random Forests and Decision Trees. *International Journal of Computer Science Issues (IJCSI), c. 9*(5), s. 272-278.

Altın Yavuz, A. ve Yavuz, H. S. (2021). Denetimli Makine Öğrenme Yöntemleri İle Yüzey Su Kalitesinin Sınıflandırılması. *Biyoloji Bilimleri Araştırma Dergisi, c. 14*(2), s. 142-155.

Arasu, B. S., Seelan, B. J. B. ve Thamaraiselvan, N. (2020). A machine learning-based approach to enhancing social media marketing. *Computers & Electrical Engineering, c. 86*, s. 106723.

Asghar, F. ve Asif, M. (2018). Salaries, performance and mediating effect of altruistic behavior: fresh statistical evidence from the national basketball association. *International Journal of Humanities. Art Soc Stud (IJHAS), c. 3*, s. 71-79.

Aszemi, N. M. ve Dominic, P. (2019). Hyperparameter Optimization in Convolutional Neural Network Using Genetic Algorithms. *International Journal of Advanced Computer Science and Applications, c. 10*(6), s. 269-278.

Atalay, M. ve Çelik, E. (2017). Büyük veri analizinde yapay zekâ ve makine öğrenmesi uygulamalari-artificial intelligence and machine learning applications in big data analysis. *Mehmet Akif Ersoy Üniversitesi Sosyal Bilimler Enstitüsü Dergisi, c. 9*(22), s. 155-172.

AYYILDIZ, E. (2018). Amerika Basketbol Ligi (NBA) maç sonuçlarının yapay sinir ağları ile tahmini. *Gaziantep Üniversitesi Spor Bilimleri Dergisi, c. 3*(1), s. 40-53.

Azevedo, A. ve Santos, M. F. (2008). KDD, SEMMA and CRISP-DM: a Parallel Overview. *IADS-DM*, s. 182-185.

Azzone, M., Barucci, E., Moncayo, G. G. ve Marazzina, D. (2022). A machine learning model for lapse prediction in life insurance contracts. *Expert Systems with Applications, c. 191*, s. 116261.

Babaoğlu, A. (2015). *Veri Madenciliği Yöntemleri Ve Bir Uygulama.* Yayımlanmamış Yüksek Lisans Tezi Yayımlanmamış Yüksek Lisans Tezi. Selçuk Üniversitesi Fen Bilimleri Enstitüsü, Konya.

Banfield, R. E., Hall, L. O., Bowyer, K. W. ve Kegelmeyer, W. P. (2006). A Comparison of Decision Tree Ensemble Creation Techniques. *IEEE transactions on pattern analysis and machine intelligence, c. 29*(1), s. 173-180.

Barut, Z. ve Bilgin, T. T. (2023). Konut Fiyatlarının Tahmini için Polinomsal Regresyon ve Yapay Sinir Ağları Yöntemlerinin Uygulamalı Karşılaştırılması. *Süleyman Demirel Üniversitesi Fen Bilimleri Enstitüsü Dergisi, c. 27*(1), s. 152-159.

Battineni, G., Sagaro, G. G., Chinatalapudi, N. ve Amenta, F. (2020). Applications Of Machine Learning Predictive Models In The Chronic Disease Diagnosis. *Journal of personalized medicine, c. 10*(2), s. 21.

Bedi, J. ve Toshniwal, D. (2019). Deep learning framework to forecast electricity demand. *Applied energy, c. 238*, s. 1312-1326.

Bergstra, J. ve Bengio, Y. (2012). Random search for hyper-parameter optimization. *Journal of machine learning research, c. 13*(2).

Biau, G. ve Scornet, E. (2016). A Random Forest Guided Tour. *Test, c. 25*, s. 197-227.

Bilen, M. (2022). *Yapay Zekânın Değıştırdığı Dınamıkler* Eğitim Yayınevi.

Bilgin, M. (2017). Gerçek Veri Setlerinde Klasik Makine Öğrenmesi Yöntemlerinin Performans Analizi. *Breast, c. 2*(9), s. 683.

Bonaccorso, G. (2017). *Machine Learning Algorithms*: Packt Publishing Ltd.

Breiman, L. (1996). Bagging Predictors. *Machine learning, c. 24*, s. 123-140.

Brownlee, J. (2018). What is the Difference Between a Batch and an Epoch in a Neural Network. *Machine Learning Mastery, c. 20*, s. 1-5.

Bülbül, M. A. (2023). Kuru Fasulye Tohumlarının Çok Sınıflı Sınıflandırılması İçin Hibrit Bir Yaklaşım. *Journal of the Institute of Science and Technology, c. 13*(1), s. 33-43.

Can, Ö. Y. (2020). *Makine öğrenmesi teknikleri kullanılarak kredi risk analizi.* Yayımlanmamış Yüksek Lisans Tezi. İstanbul Aydın Üniversitesi Lisansüstü Eğitim Enstitüsü, İstanbul.

Carbonell, J. G., Michalski, R. S. ve Mitchell, T. M. (1983). An overview of machine learning. *Machine learning*, s. 3-23.

Chen, T. ve Guestrin, C. (2016). *Xgboost: A scalable tree boosting system.* Paper presented at the Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining.

Chetan, N. B., Gong, J., Zhou, H., Bi, D., Lan, J. ve Qie, L. (2020). *An overview of recent progress of lane detection for autonomous driving.* Paper presented at the 2019 6th International conference on dependable systems and their applications (DSA).

Chun, H., Creegan, D., Majedi, O., Smolyak, D. ve Valcarcel, B. (2020). *Data-Driven Approaches to NBA Team Evaluation and Building.*

Cibaroğlu, M. O. ve Yalçınkaya, B. (2019). Belge Ve Arşiv Yönetimi Süreçlerinde Büyük Veri Analitiği Ve Yapay Zekâ Uygulamaları. *Bilgi Yönetimi, c. 2*(1), s. 44-58.

Coronato, A., Naeem, M., De Pietro, G. ve Paragliola, G. (2020). Reinforcement learning for intelligent healthcare applications: A survey. *Artificial Intelligence in Medicine, c. 109*, s. 101964.

Coşkun, F. ve Gülleroğlu, H. D. (2021). Yapay Zekânın Tarih İçindeki Gelişimi Ve Eğitimde Kullanılması. *Ankara University Journal of Faculty of Educational Sciences (JFES), c. 54*(3), s. 947-966.

Cömert, N. ve Kaymaz, M. (2019). Araç Sigortası Hilelerinde Veri Madenciliğinin Kullanımı. *Marmara Üniversitesi İktisadi ve İdari Bilimler Dergisi, c. 41*(2), s. 364-390.

Çelebi, V. ve İnal, A. (2019). Yapay Zekâ Bağlamında Etik Problemi. *Journal of International Social Research, c. 12*(66).

Çene, E. (2022). MAKİNE ÖĞRENMESİ YÖNTEMLERİYLE EUROLEAGUE BASKETBOL MAÇ SONUÇLARININ TAHMİN EDİLMESİ VE MAÇ SONUÇLARI ÜZERİNDE EN ETKİLİ DEĞİŞKENLERİN BULUNMASI. *Spor ve Performans Araştırmaları Dergisi, c. 13*(1), s. 29-52.

Çetin, Ö. O. ve Kotan, M. (2023). Yapay Zekânın Finans Alanında Kullanımı Ve Muhtemel Etkileri. *Finans, Davranışsal Finans ve Güncel Yaklaşımlar*, s. 61-69.

Daş, B. ve Türkoğlu, İ. (2014). DNA dizilimlerinin sınıflandırılmasında karar ağacı algoritmalarının karşılaştırılması.

de Saá Guerra, Y., González, J. M. M., López, N. A., Montesdeoca, S. S., Ruiz, D. R. ve Manso, J. M. G. (2011). Analysis of competitiveness in the NBA regular seasons. *Baltic Journal of Sport and Health Sciences, c. 1*(80).

Del Corral, J., García-Unanue, J. ve Herencia-Quintanar, F. (2016). Are NBA policies that promote long-term competitive balance effective? What is the price? *The Open Sports Sciences Journal, c. 9*(1), s. 81-93.

Demirci Gülpınar, V. (2022). Dijital Reklamcılıkta Makine Öğrenmesi ve Veri Gizliliği. *Kent Akademisi, c. 15*(3), s. 1455-1474.

Diaz, G. I., Fokoue-Nkoutche, A., Nannicini, G. ve Samulowitz, H. (2017). An effective algorithm for hyperparameter optimization of neural networks. *IBM Journal of Research and Development, c. 61*(4/5), s. 9: 1-9: 11.

Dubey, S. R., Singh, S. K. ve Chaudhuri, B. B. (2022). Activation functions in deep learning: A comprehensive survey and benchmark. *Neurocomputing*, s. 1-18.

Duman, E. A. (2022). *Developing a novel decision support system for performance evaluation and team formation in sports.* Yayımlanmamış Yüksek Lisans Tezi. Marmara University Institute for Graduate Studies In Pure and Applied Sciences, İstanbul.

Dündar, D. R., Sarıçiçek, İ., Çınar, E. ve Yazıcı, A. (2021). Kestirimci Bakımda Makine Öğrenmesi: Literatür Araştırması. *Eskişehir Osmangazi Üniversitesi Mühendislik ve Mimarlık Fakültesi Dergisi, c. 29*(2), s. 256-276.

Dündar, Y. (2021). *Türk Medyasında Dijitalleşen Spor Yayıncılığı:" Socrates Dergi" YouTube Kanalı Örneği.* Yayımlanmamış Yüksek Lisans Tezi. Marmara Universitesi Sosyal Bilimler Enstitüsü, İstanbul.

Erden, C. (2021). *Python İle Veri Madenciliği*: KODLAB YAYIN DAĞITIM YAZILIM LTD. ŞTİ.

Erdoğan, M. Ö. (2021). *Turkish Airlines Euroleague'de oynanan maçların maç sonucunun ve final four oynama başarısının makine öğrenmesi ile tahmin edilmesi.* Mimar Sinan Güzel Sanatlar Üniversitesi,

Ersöz, F. ve Çınar, Y. (2021). Veri madenciliği ve makine öğrenimi yaklaşımlarının karşılaştırılması: Tekstil sektöründe bir uygulama. *Avrupa Bilim ve Teknoloji Dergisi,* (29), s. 397-414.

Ertug, G. ve Castellucci, F. (2013). Getting what you need: How reputation and status affect team performance, hiring, and salaries in the NBA. *Academy of Management Journal, c. 56*(2), s. 407-431.

Fujiyoshi, H., Hirakawa, T. ve Yamashita, T. (2019). Deep learning-based image recognition for autonomous driving. *IATSS research, c. 43*(4), s. 244-252.

García-Nieto, P. J., García-Gonzalo, E. ve Paredes-Sánchez, J. P. (2021). Prediction of the critical temperature of a superconductor by using the WOA/MARS, Ridge, Lasso and Elastic-net machine learning techniques. *Neural Computing and Applications, c. 33*, s. 17131-17145.

Gedik, Y. (2022). Bir Oyundan Daha Fazlası: Spor Pazarlaması Avantajları, Zorlukları Ve Stratejileri Üzerine Kavramsal Bir Değerlendirme. *Yeni Medya Elektronik Dergisi, c. 6*(2), s. 169-187.

Georgievski, B. ve Vrtagic, S. (2021). Machine learning and the NBA Game. *Journal of Physical Education and Sport, c. 21*(6), s. 3339-3343.

Ghojogh, B. ve Crowley, M. (2019). The theory behind overfitting, cross validation, regularization, bagging, and boosting: tutorial. *arXiv preprint arXiv:1905.12787*.

Goldrosen, J. M. (2018). *The Motivations and Effects of the NBA Salary Cap.* Yayımlanmamış Proje Senior Projects. Bard College The Division of Social Studies, New York.

Golgiyaz, S. (2023). Bilgisayar Yazılımı Eğitiminde ChatGPT ChatGPT in Computer Software Education. *arXiv preprint arXiv:1709.02435*.

Gök, İ. Y., Teker, T. ve Taşan, B. (2021). NBA Takımlarının Maaş Bütçelerinin Sportif Başarıya Etkisi. *Yaşar Üniversitesi E-Dergisi, c. 16*(63), s. 1105-1118.

Gökalp, Ö. M. (2022). *Makine Öğrenmesi*.

Grasso, J. (2010). *Historical Dictionary of Basketball*: Scarecrow Press.

Güney, E. ve Yavuz, H. (2020). Yapay Zekâ İle Sanatsal Üretim Pratiğinde Sanatçının Rolü Ve Değişen Sanat Olgusu. *Sanat ve Tasarım Dergisi,* (26), s. 415-439.

Güvenç, E., Sakal, M., Çetin, G. ve Özkaraca, O. (2022). Öğrencilerin Dersteki Niteliklerinin Makine Öğrenmesi Teknikleri Kullanılarak Sınıflandırılması. *Düzce Üniversitesi Bilim ve Teknoloji Dergisi, c. 10*(3), s. 1359-1371.

Hair Jr, J. F. ve Sarstedt, M. (2021). Data, measurement, and causal inferences in machine learning: opportunities and challenges for marketing. *Journal of Marketing Theory and Practice, c. 29*(1), s. 65-77.

Han, J., Pei, J. ve Tong, H. (2022). *Data mining: concepts and techniques*: Morgan kaufmann.

Hashimoto, D. A., Rosman, G., Rus, D. ve Meireles, O. R. (2018). Artificial intelligence in surgery: promises and perils. *Annals of surgery, c. 268*(1), s. 70-76.

Hentilä, D. (2019). *The link between salary and performance: Are NBA players overpaid?* Yayımlanmamış Yüksek Lisans Tezi. Tallinn University Of Technology School of Business and Governance, Tallinn.

Hodson, T. O. (2022). Root-mean-square error (RMSE) or mean absolute error (MAE): when to use them or not. *Geoscientific Model Development, c. 15*(14), s. 5481-5487.

Horvat, T., Job, J., Logozar, R. ve Livada, Č. (2023). A Data-Driven Machine Learning Algorithm for Predicting the Outcomes of NBA Games. *Symmetry, c. 15*(4), s. 798.

İpek, S. (2019). Yapay zekanın toplum üzerindeki etkisi ve yapay zekâ (AI) filmi bağlamında yapay zekaya bakış. *Uluslararası Ders Kitapları ve Eğitim Materyalleri Dergisi, c. 2*(2), s. 203-215.

Jackson, J. (2002). Data mining; a conceptual overview. *Communications of the Association for Information Systems, c. 8*(1), s. 19.

Jain, A., Jain, S., Pancinovia, N. M. ve George, J. P. (2022). *A Non-Linear Approach to Predict the Salary of NBA Athletes using Machine Learning Technique.* Paper presented at the 2022 International Conference on Trends in Quantum Computing and Emerging Business Technologies (TQCEBT).

Janiesch, C., Zschech, P. ve Heinrich, K. (2021). Machine learning and deep learning. *Electronic Markets, c. 31*(3), s. 685-695.

Kahraman, C. (2022). *Prediction of basketball game results using machine learning algorithms: analysis of NBA & TBL.* Middle East Technical University,

Kaimakamis, C. (2021). *Sports Analytics Algorithm for NBA Champion Prediction.* Yayımlanmamış Yüksek Lisans Tezi. International Hellenic University School of Science and Technology, Thessaloniki.

Karahalilöz, O. (2021). *Öğretmen görüşlerine göre okul müdürlerinin alturistik (özgeci) liderlik davranışlarının incelenmesi.* Yayımlanmamış Yüksek Lisans Tezi. Karabük Üniversitesi Lisansüstü Eğitim Enstitüsü,

Karaoğlu, B. (2015). Makine öğrenmesi ile spor karşılaşmalarının modellenmesi. *Emo Bilimsel Dergi, c. 5*(9), s. 1-5.

Kaynar, O., Tuna, M. F., Görmez, Y. ve Deveci, M. A. (2017). Makine öğrenmesi yöntemleriyle müşteri kaybı analizi. *Cumhuriyet Üniversitesi İktisadi ve İdari Bilimler Dergisi, c. 18*(1), s. 1-14.

Kekez, I., Ćukušić, M. ve Jadrić, M. (2021). Data mining approach for business value analysis in basketball. *Zbornik Veleučilišta u Rijeci, c. 9*(1), s. 227-248.

Khan, T., Sherazi, H. H. R., Ali, M., Letchmunan, S. ve Butt, U. M. (2021). Deep learning-based growth prediction system: A use case of China agriculture. *Agronomy, c. 11*(8), s. 1551.

Kılıç Depren, S. (2019). THE EFFECTIVENESS OF DIFFERENT MACHINE LEARNING ALGORITHMS ON BASKETBALL PLAYERS'SHOOTING PERFORMANCE. *Spor ve Performans Araştırmaları Dergisi, c. 10*(3), s. 256-269.

Kıratlı, T. (2020). *Genç Basketbolcularda Modifiye Edilmiş Nba Draft Testlerinin Türkiye Normatif Değerleri* Yayımlanmamış Yüksek Lisans Tezi. Kocaeli Üniversitesi, Sağlık Bilimleri Enstitüsü, Kocaeli.

Kırca, S., Canatan, G. ve Güneş, V. (2022). Makine Öğrenmesi Yaklaşımı ile Acente Kredi Riski Hesaplama. *Bilişim Teknolojileri Dergisi, c. 15*(2), s. 159-166.

Kobat, L. R. (2018). *Win Shares & Rookie Contracts in the NBA*.

Koca, O. (2021). Veri Tabanlı Pazarlama Ve Makine Öğrenmesi İle Müşteri Bölümleme Ve Davranış Modellerinin Belirlenmesi. *Journal of Management Marketing and Logistics, c. 8*(2), s. 89-111.

Koçak, H. (2020). Çalışanların Örgütsel Bağlılıklarının Cart Karar Ağacı Algoritması İle Belirlenmesi. *Uluslararası İktisadi ve İdari Bilimler Dergisi, c. 6*(2), s. 66-87.

Kou, G., Chao, X., Peng, Y., Alsaadi, F. E. ve Herrera-Viedma, E. (2019). Machine Learning Methods For Systemic Risk Analysis İn Financial Sectors. *Technological and Economic Development of Economy, c. 25*(5), s. 716-742.

Krasheninnikova, E., García, J., Maestre, R. ve Fernández, F. (2019). Reinforcement Learning For Pricing Strategy Optimization In The Insurance Industry. *Engineering applications of artificial intelligence, c. 80*, s. 8-19.

Krizhevsky, A., Sutskever, I. ve Hinton, G. E. (2017). ImageNet classification with deep convolutional neural networks. *Communications of the ACM, c. 60*(6), s. 84-90. doi:10.1145/3065386

Kurum, V. (2020). *Spor Verisi İle Analiz Ve Tahminleme: Bir NBA Uygulaması.* Yayımlanmamış Yüksek Lisans Tezi. Bahcesehir Üniversitesi Fen Bilimleri Enstitüsü, İstanbul.

Lemmens, A. ve Croux, C. (2006). Bagging and boosting classification trees to predict churn. *Journal of Marketing Research, c. 43*(2), s. 276-286.

Li, Y., Wang, L. ve Li, F. (2021). A data-driven prediction approach for sports team performance and its application to National Basketball Association. *Omega, c. 98*, s. 102123.

Liang, H., Sun, X., Sun, Y. ve Gao, Y. (2017). Text feature extraction based on deep learning: a review. *EURASIP journal on wireless communications and networking, c. 2017*(1), s. 1-12.

Luque, A., Carrasco, A., Martín, A. ve de Las Heras, A. (2019). The impact of class imbalance in classification performance metrics based on the binary confusion matrix. *Pattern Recognition, c. 91*, s. 216-231.

Mahmood, Z., Daud, A. ve Abbasi, R. A. (2021). Using machine learning techniques for rising star prediction in basketball. *Knowledge-Based Systems, c. 211*, s. 106506.

Martínez-Plumed, F., Contreras-Ochando, L., Ferri, C., Hernández-Orallo, J., Kull, M., Lachiche, N., . . . Flach, P. (2019). CRISP-DM twenty years later: From data mining processes to data science trajectories. *IEEE Transactions on Knowledge and Data Engineering, c. 33*(8), s. 3048-3061.

McKinney, B. A., Reif, D. M., Ritchie, M. D. ve Moore, J. H. (2006). Machine Learning For Detecting Gene-Gene Interactions: A Review. *Applied bioinformatics, c. 5*, s. 77-88.

Mehonic, A., Sebastian, A., Rajendran, B., Simeone, O., Vasilaki, E. ve Kenyon, A. J. (2020). Memristors—From in-memory computing, deep learning acceleration, and spiking neural networks to the future of neuromorphic and bio-inspired computing. *Advanced Intelligent Systems, c. 2*(11), s. 2000085.

Miklosik, A., Kuchta, M., Evans, N. ve Zak, S. (2019). Towards The Adoption Of Machine Learning-Based Analytical Tools İn Digital Marketing. *Ieee Access, c. 7*, s. 85705-85718.

Mousavi, S. S., Schukat, M. ve Howley, E. (2018). *Deep reinforcement learning: an overview.* Paper presented at the Proceedings of SAI Intelligent Systems Conference (IntelliSys) 2016: Volume 2.

Murphy, R. F. (2011). An Active Role For Machine Learning In Drug Development. *Nature chemical biology, c. 7*(6), s. 327-330.

Nguyen, N. H., Nguyen, D. T. A., Ma, B. ve Hu, J. (2022). The application of machine learning and deep learning in sport: predicting NBA players' performance and popularity. *Journal of Information and Telecommunication, c. 6*(2), s. 217-235.

Ni, J., Young, T., Pandelea, V., Xue, F. ve Cambria, E. (2022). Recent advances in deep learning based dialogue systems: A systematic survey. *Artificial intelligence review*, s. 1-101.

Oğuzlar, A. (2003). Veri ön işleme. *Erciyes Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, (21).

Ötkan, C. ve Çolakoğlu, T. (2017). Salary Cap system in NBA and adaptability to Turkish basketball leaguespNBA'de uygulanan Salary Cap sistemi ve Türkiye basketbol liglerine uyarlanabilirliği. *Journal of Human Sciences, c. 14*(1), s. 963-975.

Özbalta, E., Yavuz, M. ve Kaya, T. (2022). *National Basketball Association Player Salary Prediction Using Supervised Machine Learning Methods.* Paper presented at the Intelligent and Fuzzy Techniques for Emerging Conditions and Digital Transformation: Proceedings of the INFUS 2021 Conference, held August 24-26, 2021. Volume 2.

Özdemir, E. (2020). *BASKETBOL MAÇ SONUÇLARININ MAKİNE ÖĞRENMESİ YÖNTEMLERİ İLE TAHMİN EDİLMESİ.* MUĞLA SITKI KOÇMAN ÜNİVERSİTESİ,

Özdemir, E. ve BALLI, S. (2020). TÜRKİYE ERKEKLER BASKETBOL LİGİ MAÇ SONUÇLARININ MAKİNE ÖĞRENMESİ YÖNTEMLERİ İLE TAHMİNİ. *Mühendislik Bilimleri ve Tasarım Dergisi, c. 8*(3), s. 740-752.

Özen, N. S., Saraç, S. ve Koyuncu, M. (2021). COVID-19 vakalarının makine öğrenmesi algoritmaları ile tahmini: Amerika Birleşik Devletleri örneği. *Avrupa Bilim ve Teknoloji Dergisi*, (22), s. 134-139.

Öztemel, E. (2003). Yapay sinir ağlari. *Papatya Yayincilik, Istanbul*.

Papadaki, I. ve Tsagris, M. (2020). Estimating NBA players salary share according to their performance on court: A machine learning approach. *arXiv preprint arXiv:2007.14694*, s. 1-19.

Patil, P. S. ve Dharwadkar, N. V. (2017). *Analysis Of Banking Data Using Machine Learning.* Paper presented at the 2017 International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud)(I-SMAC).

Patrot, A., Harish, H., Shambbavi, B. ve Geetha, P. (2023). *NBA Game Prediction Using Machine Learning Algorithm.* Paper presented at the 2023 International Conference on Recent Trends in Electronics and Communication (ICRTEC).

Pirim, H. (2006). Yapay zeka. *Yaşar Üniversitesi E-Dergisi, c. 1*(1), s. 81-93.

Preetha, S., Afrid, I. ve Nishchay, S. (2020). Machine Learning for Handwriting Recognition. *International Journal of Computer (IJC), c. 38*(1), s. 93-101.

Priyadarshini, I., Sahu, S., Kumar, R. ve Taniar, D. (2022). A machine-learning ensemble model for predicting energy consumption in smart homes. *Internet of Things, c. 20*, s. 100636.

Ray, S. (2019). *A quick review of machine learning algorithms.* Paper presented at the 2019 International conference on machine learning, big data, cloud and parallel computing (COMITCon).

Rong, S. ve Bao-Wen, Z. (2018). *The research of regression model in machine learning field.* Paper presented at the MATEC Web of Conferences.

Ruder, S., Peters, M. E., Swayamdipta, S. ve Wolf, T. (2019). *Transfer learning in natural language processing.* Paper presented at the Proceedings of the 2019 conference of the North American chapter of the association for computational linguistics: Tutorials.

Sağ, R. ve Hasırcı Tuğcu, Z. (2022). Akıllı Şebeke Uygulamalarında Derin Öğrenme Tekniklerinin Kullanımına İlişkin Kısa Bir İnceleme. *Emo Bilimsel Dergi, c. 13*(1), s. 41-61.

Salay, R., Queiroz, R. ve Czarnecki, K. (2017). An analysis of ISO 26262: Using machine learning safely in automotive software. *arXiv preprint arXiv:1709.02435*, s. 1-6.

Sargent, D. J. (2001). Comparison of artificial neural networks with other statistical approaches: results from medical data sets. *Cancer: Interdisciplinary International Journal of the American Cancer Society, c. 91*(S8), s. 1636-1642.

Schröer, C., Kruse, F. ve Gómez, J. M. (2021). A systematic literature review on applying CRISP-DM process model. *Procedia Computer Science, c. 181*, s. 526-534.

Sharma, S., Bhatt, M. ve Sharma, P. (2020). *Face recognition system using machine learning algorithm.* Paper presented at the 2020 5th International Conference on Communication and Electronics Systems (ICCES).

Sigler, K. ve Compton, W. (2018). NBA Players' Pay and Performance: What Counts? *Sport Journal*.

Sigler, K. J. ve Sackley, W. H. (2000). NBA players: are they paid for performance? *Managerial finance, c. 26*(7), s. 46-51.

Smithson, S. C., Yang, G., Gross, W. J. ve Meyer, B. H. (2016). *Neural networks designing neural networks: multi-objective hyper-parameter optimization.* Paper presented at the 2016 IEEE/ACM International Conference on Computer-Aided Design (ICCAD).

Somvanshi, M., Chavan, P., Tambade, S. ve Shinde, S. (2016). *A review of machine learning techniques using decision tree and support vector machine.* Paper presented at the 2016 international conference on computing communication control and automation (ICCUBEA).

Stanek, T. (2016). *Player performance and team revenues: NBA player salary analysis.* Yayımlanmamış Yüksek Lisans Tezi. Claremont McKenna College,

Şahin, Ö. E., Ayvaz, S. ve Çalımfidan, E. (2020). Sigorta Sektöründe Sahte Hasarların Tahmini İçin Geliştirilen Makine Öğrenmesi Modellerinin Kıyaslanması. *Bilişim Teknolojileri Dergisi, c. 13*(4), s. 479-489.

Şeker, A., Diri, B. ve Balık, H. H. (2017). Derin öğrenme yöntemleri ve uygulamaları hakkında bir inceleme. *Gazi Mühendislik Bilimleri Dergisi, c. 3*(3), s. 47-64.

Şeker, Ş. E. (2018). CRISP-DM: Endüstriler Arası Standart İşleme–Veri Madenciliği için (Cross Industry Standard Processing–Data Mining). *YBS Ansiklopedi, c. 5*(2), s. 10-16.

Şen, Z. (2004). *Yapay Sinir Ağları İlkeleri*: Su Vakfı.

Taşkın, Ç. ve Gökay, E. G. (2005). Veri madenciliğinde karar ağaçları ve bir satış analizi uygulaması. *Eskişehir Osmangazi Üniversitesi Sosyal Bilimler Dergisi, c. 6*(2), s. 221-239.

Teket, O. M. (2021). *Basketbol oyuncu tanımlama ve skor tespiti.* TOBB ETÜ,

Thabtah, F., Zhang, L. ve Abdelhamid, N. (2019). NBA game result prediction using feature analysis and machine learning. *Annals of Data Science, c. 6*(1), s. 103-116.

Wale, N. (2011). Machine Learning In Drug Discovery And Development. *Drug Development Research, c. 72*(1), s. 112-119.

Wang, J. ve Fan, Q. (2021). *Application of machine learning on nba data sets.* Paper presented at the Journal of Physics: Conference Series.

Wang, S. ve Summers, R. M. (2012). Machine Learning And Radiology. *Medical image analysis, c. 16*(5), s. 933-951.

Wirth, R. ve Hipp, J. (2000). *CRISP-DM: Towards a standard process model for data mining.* Paper presented at the Proceedings of the 4th international conference on the practical applications of knowledge discovery and data mining.

Wu, W. (2020). Injury Analysis Based on Machine Learning in NBA Data. *Journal of Data Analysis and Information Processing, c. 8*(4), s. 295-308.

Wu, W., Feng, K., Li, R., Sengupta, K. ve Cheng, A. (2020). Classification of NBA salaries through player statistics. *University of California-Berkeley*, s. 1-21.

Xiong, R., Greene, M., Tanielian, V. ve Ulibarri, J. (2017). *Research on the relationship between salary and performance of professional basketball team (nba).* Paper presented at the Proceedings of the 8th International Conference on E-business, Management and Economics.

Yang, M., Wei, Y., Liang, L., Ding, J. ve Wang, X. (2021). Performance evaluation of NBA teams: A non-homogeneous DEA approach. *Journal of the Operational Research Society, c. 72*(6), s. 1403-1414.

Yığman, C. (2018). *Predictive analysis of successful basketball shots?: The Euroleague case.* Sabancı University,

Yılmaz, A. (2021). *Yapay Zeka*: Kodlab Yayın Dağıtım Yazılım Ltd. Şti.

Yücel, A. (2021). Tüketici Yorumları Üzerine Bir Metin Madenciliği Ve Veri Boyutu İndirgeme Yaklaşımı. *Journal of Intelligent Systems: Theory and Applications, c. 4*(1), s. 8-16.

Zeiler, M. D. (2012). Adadelta: an adaptive learning rate method. *arXiv preprint arXiv:1212.5701*.

İnternet Kaynakları:

Alay, S., (2019). "Linear Regression (Ridge ve Lasso Fonksiyonları)". https://www.kaggle.com/code/colecola/linear-regression-ridge-ve-lasso-fonksiyonlar, (Erişim Tarihi: 09.05.2023).

Basketball Reference, (2023a). " Basketball Stats and History". https://www.basketball-reference.com/, (Erişim Tarihi: 14.04.2023).

Basketball Reference, (2023b). "NBA Win Shares". https://www.basketball-reference.com/about/ws.html, (Erişim Tarihi: 07.09.2023).

Bilen, B., (2021). "Karışıklık Matrisi (Confusion Matrix)". https://burhanbilen.medium.com/kar%C4%B1%C5%9F%C4%B1kl%C4%B1k-matrisi-confusion-matrix-990dfc718653, (Erişim Tarihi: 10.05.2023).

Çınar, U. K., (2018). "Yapay Sinir Ağları". https://www.veribilimiokulu.com/yapay-sinir-aglari/, (Erişim Tarihi: 08.05.2023).

Eurohoops.net, (2022). "James Harden, Maaşında İndirime Gitmesine Rağmen Hakkının Verilmediğini Düşünüyor". https://www.eurohoops.net/tr/nba-news-tr/1402613/james-harden-maasinda-indirime-gitmesine-ragmen-hakkinin-verilmedigini-dusunuyor/, (Erişim Tarihi: 07.09.2023).

Hoopshype, (2023). " Hoopshype". https://hoopshype.com, (Erişim Tarihi: 14.04.2023).

Knime, (2023a). "Knime". https://www.knime.com/, (Erişim Tarihi: 02.03.2023).

Knime, (2023b). “Getting Started Guide”. https://www.knime.com/early-access-knime-ap-v5-getting-started, (Erişim Tarihi: 02.03.2023).

López, F., (2021). "Ensemble Learning: Bagging & Boosting". https://towardsdatascience.com/ensemble-learning-bagging-boosting-3098079e5422, (Erişim Tarihi: 09.05.2023).

Nagpal, A., (2017). "Decision Tree Ensembles- Bagging and Boosting". https://towardsdatascience.com/decision-tree-ensembles-bagging-and-boosting-266a8ba60fd9, (Erişim Tarihi: 09.05.2023).

NBA Hoops Online, (2023). "NBA History". https://nbahoopsonline.com/History/, (Erişim Tarihi: 14.04.2023).

NBA Media Ventures, (2023a). "All Teams". https://www.nba.com/teams (Erişim Tarihi: 18.04.2023).

NBA Media Ventures, (2023b). "NBA Advanced Stats Franchise History". https://www.nba.com/stats/history (Erişim Tarihi: 14.04.2023).

NBA Media Ventures, (2023c). "NBA Draft Lottery: Odds, history and how it Works". https://www.nba.com/news/nba-draft-lottery-explainer, (Erişim Tarihi: 24.04.2023).

NBA Media Ventures, (2023d). "NBA salary cap and tax level expected to rise in 2023-24". https://www.nba.com/news/nba-salary-cap-and-tax-level-expected-to-rise-in-2023-24 (Erişim Tarihi: 18.04.2023).

NBA, (2023). " NBA". https://www.nba.com/ , (Erişim Tarihi: 14.04.2023).

Sandomir, R., (2014). "N.B.A. Is Said to Continue Network Deals". https://www.nytimes.com/2014/10/06/sports/basketball/nba-said-to-be-near-new-tv-deal-for-24-billion.html (Erişim Tarihi: 25.04.2023).

Scope Medya A.Ş, (2016). "Shaq O'Neal; nam-ı diğer "Win Diesel"". https://medyascope.tv/2016/03/08/shaq-nam-i-diger-win-diesel/, (Erişim Tarihi: 07.09.2023).

Silipo, R., (2019). "From a Single Decision Tree to a Random Forest". https://towardsdatascience.com/from-a-single-decision-tree-to-a-random-forest-b9523be65147, (Erişim Tarihi: 08.05.2023).

Slatery, K., (2020). "Decision Trees: Understanding the Basis of Ensemble Methods". https://towardsdatascience.com/decision-trees-understanding-the-basis-of-ensemble-methods-e075d5bfa704, (Erişim Tarihi: 09.05.2023).

Spotrac, (2023). " Spotrac". https://www.spotrac.com, (Erişim Tarihi: 14.04.2023).

EK 1- NBA Takımları ve Kısaltmaları

Takımlar	Kısaltmaları
Phoenix Suns	PHX-PHO
Memphis Grizzlies	MEM
Golden State Warriors	GSW
Miami Heat	MIA
Dallas Mavericks	DAL
Boston Celtics	BOS
Milwaukee Bucks	MIL
Philadelphia 76ers	PHI
Utah Jazz	UTA
Denver Nuggets	DEN
Toronto Raptors	TOR
Chicago Bulls	CHI
Minnesota Timberwolves	MIN
Brooklyn Nets	BKN
Cleveland Cavaliers	CLE
Atlanta Hawks	ATL
Charlotte Hornets	CHA
LA Clippers	LAC
New York Knicks	NYK
New Orleans Pelicans	NOP
Washington Wizards	WAS
San Antonio Spurs	SAS
Los Angeles Lakers	LAL
Sacramento Kings	SAC
Portland Trail Blazers	POR
Indiana Pacers	IND
Oklahoma City Thunder	OKC
Detroit Pistons	DET
Orlando Magic	ORL
Houston Rockets	HOU

EK 2- Veri Setindeki Değişkenler ve Anlamları

Değişkenler	Anlamları
ID	Kimlik Numarası
Player	Oyuncu
Team	Takım
Team-Abb	Takım Kısaltması
Last Season Win	Son Sezon Kazanma
Playoffs katılma	Playofflara Katılma
Playoffs Win	Playofflarda Kazanma
Total Cap 21-22	2021-2022 Sezonu Toplam Maaş Tavanı
Cap Space 21-22	2021-2022 Sezonu Maaş Boşluğu
Total Cap 22-23	2022-2023 Sezonu Toplam Maaş Tavanı
Cap Space 22-23	2022-2023 Sezonu Maaş Boşluğu
InstagramTeam	Takımın Instagram Hesabı
TwitterTeam	Takımın Twitter Hesabı
Age	Yaş
Pos	Pozisyon
G	Önceki Sezon Oynadığı Maç Sayısı
GS	Önceki Sezon Başlangıç Oyuncusu Olunan Önceki Sezon Maç Sayısı
MP	Önceki Sezon Maç Başına Oynama Süresi (Dakika)
FG	Önceki Sezon Saha İçi İsabet
FGA	Önceki Sezon Saha İçi Denemeleri (Ortalama)
FG%	Önceki Sezon Saha İçi İsabet Yüzdesi
3P	Önceki Sezon Üç Sayılık Atış İsabet Sayısı
3PA	Önceki Sezon Üç Sayılık Atış Deneme Sayısı
3P%	Önceki Sezon Üç Sayılık Atış Yüzdesi
2P	Önceki Sezon İki Sayılık Atış Sayısı
2PA	Önceki Sezon İki Sayılık Atış Deneme Sayısı
2P%	Önceki Sezon İki Sayılık Atış Yüzdesi
eFG%	Önceki Sezon Etkin Saha İçi İsabet
FT	Önceki Sezon Serbest Atış Sayısı
FTA	Önceki Sezon Serbest Atış Deneme Sayısı
FT%	Önceki Sezon Serbest Atış Yüzdesi
ORB	Önceki Sezon Hücum Ribaunt
DRB	Önceki Sezon Defansif/Savunma Ribaunt
TRB	Önceki Sezon Toplam Ribaunt
AST	Önceki Sezon Asist
STL	Önceki Sezon Top Çalma
BLK	Önceki Sezon Blok

TOV	Önceki Sezon Top Kaybı
PF	Önceki Sezon Kişisel Faul
PTS	Önceki Sezon Maç Başına Attığı Ortalama Sayı
G(1)	Kariyerinde Oynadığı Maç Sayısı
GS(1)	Kariyerinde Başlangıç Oyuncusu Olunan Maç Sayısı
MP(1)	Kariyerinde Maç Başına Oynama Süresi (Dakika)
FG(1)	Kariyerinde Saha İçi İsabet
FGA(1)	Kariyerinde Saha İçi Denemeleri (Ortalama)
FG%(1)	Kariyerinde Saha İçi İsabet Yüzdesi
3P(1)	Kariyerinde Üç Sayılık Atış İsabet Sayısı
3PA(1)	Kariyerinde Üç Sayılık Atış Deneme Sayısı
3P%(1)	Kariyerinde Üç Sayılık Atış Yüzdesi
2P(1)	Kariyerinde İki Sayılık Atış Sayısı
2PA(1)	Kariyerinde İki Sayılık Atış Deneme Sayısı
2P%(1)	Kariyerinde İki Sayılık Atış Yüzdesi
eFG%(1)	Kariyerinde Etkin Saha İçi İsabet
FT(1)	Kariyerinde Serbest Atış Sayısı
FTA(1)	Kariyerinde Serbest Atış Deneme Sayısı
FT%(1)	Kariyerinde Serbest Atış Yüzdesi
ORB(1)	Kariyerinde Hücum Ribaunt
DRB(1)	Kariyerinde Defansif/Savunma Ribaunt
TRB(1)	Kariyerinde Toplam Ribaunt
AST(1)	Kariyerinde Asist
STL(1)	Kariyerinde Top Çalma
BLK(1)	Kariyerinde Blok
TOV(1)	Kariyerinde Top Kaybı
PF(1)	Kariyerinde Kişisel Faul
PTS(1)	Kariyerinde Maç Başına Attığı Ortalama Sayı
G(2)	Playoff Oynadığı Maç Sayısı
GS(2)	Playoff Başlangıç Oyuncusu Olunan Maç Sayısı
MP(2)	Playoff Maç Başına Oynama Süresi (Dakika)
FG(2)	Playoff Saha İçi İsabet
FGA(2)	Playoff Saha İçi Denemeleri (Ortalama)
FG%(2)	Playoff Saha İçi İsabet Yüzdesi
3P(2)	Playoff Üç Sayılık Atış İsabet Sayısı
3PA(2)	Playoff Üç Sayılık Atış Deneme Sayısı
3P%(2)	Playoff Üç Sayılık Atış Yüzdesi
2P(2)	Playoff İki Sayılık Atış Sayısı
2PA(2)	Playoff İki Sayılık Atış Deneme Sayısı
2P%(2)	Playoff İki Sayılık Atış Yüzdesi
eFG%(2)	Playoff Etkin Saha İçi İsabet

FT(2)	Playoff Serbest Atış Sayısı
FTA(2)	Playoff Serbest Atış Deneme Sayısı
FT%(2)	Playoff Serbest Atış Yüzdesi
ORB(2)	Playoff Hücum Ribaunt
DRB(2)	Playoff Defansif/Savunma Ribaunt
TRB(2)	Playoff Toplam Ribaunt
AST(2)	Playoff Asist
STL(2)	Playoff Top Çalma
BLK(2)	Playoff Blok
TOV(2)	Playoff Top Kaybı
PF(2)	Playoff Kişisel Faul
PTS(2)	Playoff Maç Başına Attığı Ortalama Sayı
INSTAGRAM	Oyuncunun Instagram Hesabı
TWITTER	Oyuncunun Twitter Hesabı
Experience	Oyuncunun Tecrübesi
Salary_21-22	2021-2022 Sezonu Maaşı
Salary_21-22_Current	2021-2022 Sezonu Güncel Maaşı
Salary_22-23	2022-2023 Sezonu Maaşı
WS_21-22	2021-2022 Sezonu Win Share
WS_Career	Kariyer Boyunca Win Share

ÖZGEÇMİŞ

Adı ve SOYADI	Tuba SARI
EĞİTİM DURUMU	
Lisans Diploması	Akdeniz Üniversitesi Yönetim Bilişim Sistemleri
Yüksek Lisans Diploması	Akdeniz Üniversitesi Yönetim Bilişim Sistemleri Tezli YL
Tez Konusu	Makine Öğrenmesi Yöntemleri ile NBA Oyuncu Değeri Öncülerinin Tahmin Edilmesi
Yabancı Dil / Diller	İngilizce, Almanca
BİLİMSEL FAALİYETLER	
İŞ DENEYİMİ	2021-2023 İzmir Demokrasi Üniversitesi / İktisadi ve İdari Bilimler Fakültesi / Yönetim Bilişim Sistemleri Bölümü - Araştırma Görevlisi 2023-Halen Akdeniz Üniversitesi / Uygulamalı Bilimler Fakültesi / Yönetim Bilişim Sistemleri Bölümü -Araştırma Görevlisi