T.C.

YILDIZ TEKNİK ÜNİVERSİTESİ

FEN BİLİMLERİ ENSTİTÜSÜ

MAKİNE ÖĞRENMESİ YÖNTEMLERİ KULLANILARAK MAÇ ÖNCESİ VE MAÇ ESNASINDA FUTBOL ANALİZİ

Ozan BARLAS

YÜKSEK LİSANS TEZİ

Endüstri Mühendisliği Anabilim Dalı

Endüstri Mühendisliği Programı

Danışman

Dr. Öğr. Üyesi Tufan DEMİREL

Mayıs, 2020

T.C.

YILDIZ TEKNİK ÜNİVERSİTESİ

FEN BİLİMLERİ ENSTİTÜSÜ

MAKİNE ÖĞRENMESİ YÖNTEMLERİ KULLANILARAK MAÇ ÖNCESİ VE MAÇ ESNASINDA FUTBOL ANALİZİ

Ozan BARLAS tarafından hazırlanan tez çalışması 14.07.2020 tarihinde aşağıdaki jüri tarafından Yıldız Teknik Üniversitesi Fen Bilimleri Enstitüsü Endüstri Mühendisliği Anabilim Dalı, Endüstri Mühendisliği Programı **YÜKSEK LİSANS TEZİ** olarak kabul edilmiştir.

Dr. Öğr. Üyesi Tufan DEMİREL
Yıldız Teknik Üniversitesi
Danışman

Jüri Üyeleri

Dr. Öğr. Üyesi Tufan DEMİREL
Yıldız Teknik Üniversitesi _____

Prof. Dr. Semih ÖNÜT
Yıldız Teknik Üniversitesi _____

Doç. Dr. Muhammet DEVECİ
Deniz Harp Okulu _____

Danışmanım Dr. Öğr. Üyesi Tufan DEMİREL sorumluluğunda tarafımca hazırlanan Makine Öğrenmesi Yöntemleri Kullanılarak Maç Öncesi Ve Maç Esnasında Futbol Analizi başlıklı çalışmada veri toplama ve veri kullanımında gerekli yasal izinleri aldığımı, diğer kaynaklardan aldığım bilgileri ana metin ve referanslarda eksiksiz gösterdiğimi, araştırma verilerine ve sonuçlarına ilişkin çarpıtma ve/veya sahtecilik yapmadığımı, çalışmam süresince bilimsel araştırma ve etik ilkelerine uygun davrandığımı beyan ederim. Beyanımın aksinin ispatı halinde her türlü yasal sonucu kabul ederim.

Ozan BARLAS

İmza

Bu çalışma, Yıldız Teknik Üniversitesi Bilimsel Araştırma Proje Koordinatörlüğü' nün numaralı projesi ile desteklenmiştir.

Aileme

ve

sevdiklerime

TEŞEKKÜR

Çalışmanın her anında bana yol gösteren ve bilgilerini, deneyimini aktaran değerli hocam Dr. Öğr. Üyesi Tufan DEMİREL'e teşekkür ederim.

Ozan BARLAS

İÇİNDEKİLER

ŞEKİL LİSTESİ x

TABLO LİSTESİ xiii

ÖZET xiv

ABSTRACT xv

1 GİRİŞ 1

1.1 Literatür Özeti 1
1.2 Tezin Amacı 3
1.3 Hipotez 4

2 MAKİNE ÖĞRENMESİ 5

Makine Öğrenmesinin Kullanım Alanları 5
Öğrenme Yöntemine Göre Makine Öğrenmesi Algoritmaları 6
Çalışma Biçimlerine Göre Algoritmalar 8

3 YAPAY SİNİR AĞLARI 10

Yapay Sinir Ağları Kullanım Alanları 10
Yapay Sinir Ağları Özellikleri 10
Yapay Sinir Yapısı 11
Yapay Sinir Ağları İle Geliştirilen Modeller 12

4 FUTBOL MODELİ KURGULANMASI 13

4.1 Parametre Seçimi 13
4.2 Veri Setinin Model Kurgusundan Önce İşlenmesi 21
4.3 Hedef Belirlenmesi 22
4.4 Eğitim Ve Test Verisi Belirlenmesi 22

5 TAHMİNLEME MODELİ 26

5.1 Lojistik Regresyon 26
5.2 Lojistik Regresyon Sonuç & Validasyon 26
5.3 Neural Network (Sinir Ağları) 28
5.4 Neural Network (Sinir Ağları) Sonuç & Validasyon 29

6 İDDAA MODELİ 32

6.1 Lojistik Regresyon & Başarı Oranları 32
6.2 Neural Network (Sinir Ağları) & Başarı Oranları 33
7 GÜRÜLTÜ AZALTMA **35**
7.1 Lojistik Regresyon & Başarı Oranları 39
7.2 Neural Network (Sinir Ağları) & Başarı Oranları 39
8 İKİNCİ MODEL **41**
Deney 1 (Uzun Dönem Katsayı) 41
Deney 2 (V9-V12 - Tanımsız Değer Ataması) 44
Deney 3 (Kısa Dönem Son X - Karşılıklı Gol) 45
9 MODEL PARAMETRİZASYONU **49**
9.1 Liglerin Sınıflandırılması 49
9.2 Hafta Numarasına Göre Sınıflandırma 50
9.3 Neural Network (Sinir Ağları) Modeli 51
9.4 Modelin Öğrenilmesi Ve Saklanması İlk Aşama 52
9.5 Modelin Saklanma Altyapısı 53
10 CANLI SKOR MODELİ (MAÇ İÇİ VERİLER) **54**
10.1 Parametreler 54
10.2 Yöntem 56
10.3 Dakika Bazında Tahminleme Eğilimleri 66
10.4 Dakika Bazında Tahminleme Performansları 68
11 CANLI SKOR MODELİ (4 ARALIKLI AYRIŞTIRMA) **70**
11.1 Dakika Bazında Tahminleme Eğilimleri 72
11.2 Dakika Bazında Tahminleme Performansları 74
12 CANLI SKOR MODELİ (6 ARALIKLI AYRIŞTIRMA) **76**
12.1 Dakika Bazında Tahminleme Eğilimleri 77
12.2 Dakika Bazında Tahminleme Performansları 79
13 CANLI SKOR MODELİ (MAÇ SONU PARAMETRE SAYISINI ARTTIRMA) **82**
13.1 Dakika Bazında Tahminleme Eğilimleri 83
13.2 Dakika Bazında Tahminleme Performansları 84
14 BULGULAR VE TARTIŞMA **86**
15 SONUÇ VE ÖNERİLER **89**
KAYNAKÇA **91**

TEZDEN ÜRETİLMİŞ YAYINLAR **92**

KISALTMA LİSTESİ

ANN	Yapay Sinir Ağları
LR	Lojistik Regresyon
LTR	Uzun Dönem Katsayı
STR	Kısa Dönem Katsayı

ŞEKİL LİSTESİ

Şekil 4. 1 İlk model kurgulama yöntemi 17
Şekil 4. 2 Python ile skorlama yöntemi 17
Şekil 4. 3 Skorlama işlemi sonuçları 18
Şekil 4. 4 Pearson korelasyon sonuçları 19
Şekil 4. 5 Sperman korelasyon sonuçları 20
Şekil 4. 6 Python hedef seçimi 22
Şekil 4. 7 Python eğitim verisi seçimi 25
Şekil 4. 8 Python test verisi seçimi 25
Şekil 5. 1 Python tahminleme modeli lr yöntemi 26
Şekil 5. 2 Python tahminleme modeli lr hiperparametre 26
Şekil 5. 3 Tahminleme modeli lr sonuçları 27
Şekil 5. 4 Tahminleme modeli lr hata matrisi 28
Şekil 5. 5 Tahminleme modeli ann hiperparametre seçimi 29
Şekil 5. 6 Tahminleme modeli ann sonuçları 29
Şekil 5. 7 Tahminleme modeli ann hata matrisi 30
Şekil 5. 8 Tahminleme modeli ann sonuçları - 2 31
Şekil 6. 1 İddaa modeli parametre seçimi 32
Şekil 6. 2 İddaa modeli lr hata matrisi 33
Şekil 6. 3 İddaa modeli ann hiperparametre seçimi 33
Şekil 6. 4 İddaa modeli ann sonuçları 34
Şekil 6. 5 İddaa modeli ann hata matrisi 34
Şekil 7. 1 V16 – v17 parametreleri saçılım grafiği 36
Şekil 7. 2 Büyük lig seçimi 36
Şekil 7. 3 V16 – v17 parametreleri saçılım grafiği - 2 37
Şekil 7. 4 Büyük lig modeli parametre seçimi 37
Şekil 7. 5 Büyük lig modeli skorlama 38
Şekil 7. 6 Büyük lig modeli cross validation (çapraz doğrulama) sonuçları 38
Şekil 7. 7 Büyük lig modeli lr hata matrisi 39
Şekil 7. 8 Büyük lig modeli ann hiperparametre seçimi 40
Şekil 7. 9 Büyük lig modeli ann hata matrisi 40
Şekil 8. 1 Büyük lig modeli ltr saçılımı 42

Şekil 8. 2 Büyük lig modeli ltr dağılımı 42
Şekil 8. 3 Büyük lig ltr modeli skorlama 43
Şekil 8. 4 Büyük lig ltr modeli parametre seçimi 43
Şekil 8. 5 Büyük lig ltr modeli sonuçları 44
Şekil 8. 6 Büyük lig tanımsız değer modeli sonuçları 45
Şekil 8. 7 Büyük lig kısa dönem modeli skorlama 47
Şekil 8. 8 Büyük lig kısa dönem modeli sonuçları 48
Şekil 9. 1 Geçmiş ve aktif sezon parametreleri 51
Şekil 9. 2 Büyük ve küçük lig parametreleri 52
Şekil 9. 3 Modelin öğrenilmesi aşaması 53
Şekil 10. 1 Maç içi model kurgusu 64
Şekil 10. 2 Maç içi model parametreleri 65
Şekil 10. 3 Canlı skor modeli sonuçları 66
Şekil 10. 4 LR dakika bazında tahminleme eğilimleri 67
Şekil 10. 5 Karar verme ağacı dakika bazında tahminleme eğilimleri 67
Şekil 10. 6 Lr dakika bazında tahminleme performansları 68
Şekil 10. 7 Karar verme ağacı dakika bazında tahminleme performansları 68
Şekil 11. 1 4 aralıklı ayrıştırma 71
Şekil 11. 2 4 aralıklı ayrıştırma parametre seçimi 71
Şekil 11. 3 4 aralıklı ayrıştırma sonuçları 72
Şekil 11. 4 4 aralıklı ayrıştırma lr tahminleme eğilimleri 73
Şekil 11. 5 4 aralıklı ayrıştırma karar verme ağacı tahminleme eğilimleri 73
Şekil 11. 6 4 aralıklı ayrıştırma tahminleme eğilimleri karşılaştırması 74
Şekil 11. 7 4 aralıklı ayrıştırma lr tahminleme performansları 74
Şekil 11. 8 4 aralıklı ayrıştırma karar verme ağacı tahminleme performansları ... 75
Şekil 11. 9 4 aralıklı ayrıştırma lr tahminleme performansları karşılaştırması 75
Şekil 12. 1 6 aralıklı ayrıştırma 76
Şekil 12. 2 6 aralıklı ayrıştırma sonuçları 77
Şekil 12. 3 6 aralıklı ayrıştırma lr tahminleme eğilimleri 78
Şekil 12. 4 6 aralıklı ayrıştırma karar verme ağacı tahminleme eğilimleri 78
Şekil 12. 5 6 aralıklı ayrıştırma tahminleme eğilimleri karşılaştırması 79
Şekil 12. 6 6 aralıklı ayrıştırma lr tahminleme performansları 79
Şekil 12. 7 6 aralıklı ayrıştırma karar verme ağacı tahminleme performansları ... 80

Şekil 12. 8 6 aralıklı ayrıştırma lr tahminleme performansları karşılaştırması 80
Şekil 13. 1 Canlı skor parametre sayısını arttırma 82
Şekil 13. 2 Parametre sayısını arttırma sonuçları..... 82
Şekil 13. 3 Parametre sayısını arttırma lr tahminleme eğilimleri 83
Şekil 13. 4 Parametre sayısını arttırma karar verme ağacı tahminlemeleri 84
Şekil 13. 5 Parametre sayısını arttırma tahminleme eğilimleri karşılaştırması 84
Şekil 13. 6 Parametre sayısını arttırma lr tahminleme performansları..... 84
Şekil 13. 7 Parametre sayısını arttırma kva tahminleme performanslar 85
Şekil 13. 8 Parametre sayısını arttırma lr tahminleme performansları..... 85

TABLO LİSTESİ

Tablo 4. 1 Maç sonu parametreleri 14
Tablo 4. 2 Maç içi parametreler 15
Tablo 10. 1 Canlı skor modeli parametreler 55
Tablo 10. 2 Canlı skor modeli parametreleri min max değerleri 57
Tablo 10. 3 Parametre listesi 59
Tablo 14. 1 Deney takip çizelgesi 87

ÖZET

Makine Öğrenmesi Yöntemleri Kullanılarak Maç Öncesi Ve Maç Esnasında Futbol Analizi

Ozan BARLAS

Endüstri Mühendisliği Anabilim Dalı

Yüksek Lisans Tezi

Danışman: Dr. Öğr. Üyesi Tufan DEMİREL

İş dünyasında birçok alanda doğru veri analizi ile tahminleme gerçekleştirmek en önemli meydan okumalardan biridir. Bir spor kulübünün teknik direktörü için, farklı değişkenlere yönelik senaryo analizleri yapmak ve buna uygun takım kadrosunu sürmek gerekmektedir.

Bu çalışmada, bilimsel bakış açısıyla, kullanıcı dostu bir arayüzle başarılı olasılık tahminlemesi yapılması konusundaki ihtiyaca cevap vermek istenmiştir. Bunun için, Yapay Sniri Ağları, Lojistik Regresyon ve Karar Verme Ağacı gibi Makine Öğrenmesi prensiplerini uygulayan, spor, finans, tarım, ulaşım, enerji, savunma gibi birçok sektördeki kurumsal firmaların ve bireysel müşterilerin kullanabileceği bir olasılık hesaplayıcı ve simülatör projesi hedeflenmiştir.

Anahtar Kelimeler: Makine öğrenmesi, yapay sinir ağları, futbol analitikleri

ABSTRACT

Football Match Analysis Using Machine Learning Methods Before And During Match

Ozan BARLAS

Department of Industrial Engineering

Master of Science Thesis

Advisor: Assist. Prof. Dr. Tufan DEMİREL

Forecasting is one of the most important challenges in many areas in the business world. It is necessary to make scenario analyzes for different variables and to apply for the technical director of a sports club.

In this study, it is aimed to answer the need for successful probability estimation with a user-friendly interface from a scientific perspective. For this purpose, a probability calculator and simulator project, which implements Machine Learning principles such as Logistic Regression, Artificial Neural Network, Decision Tree which can be used by corporate companies and individual customers in many sectors such as sports, finance, agriculture, transportation, energy, defense, is aimed.

Keywords: Machine learning, artificial intelligence, football analytics

1
GİRİŞ

Günümüzde makine öğrenmesi yöntemleri bir çok farklı alanda tahminleme amacıyla kullanılmaktadır. Bu projede, makine öğrenmesi yöntemlerinin sırasıyla henüz oynanmamış futbol müsabakalarının analizi ve maç esnasında durum analizi yapılması amaçlanmıştır. Günümüzde makine öğrenmesi yöntemleri bir çok farklı alanda tahminleme amacıyla kullanılmaktadır.

1.1 Literatür Özeti

Futbol başta olmak üzere spor müsabakaları analizi için makine öğrenmesi yöntemleri kullanımının çok sayıda akademik örneği mevcuttur.

Spor etkinliklerinin sonuçlarının tahminlenmesi ile ilgili yöntemleri araştıran bir çalışmada, at yarışlarında ve çeşitli takım sporlarında bir takım tahmin konuları gündeme getirilmiştir. Yöntemler öncelikle yapılan tahmin türününe göre ayrıştırılmış, farklı değerlendirme prosedürleri incelenmiş ve alternatif tahmin yöntemleri (modeller, uzmanlar ve piyasa) karşılaştırılmıştır. Makalede ayrıca tahminlerde önyargıların varlığına ilişkin kanıtlar incelenmiştir (Stekler vd. (2010)).

Basketbol müsabakalarının tahminlenmesiyle ilgili olarak yapılan bir çalışmada, possession-based Markov model kullanılmıştır. Çalışmada geçmiş matris oluşturulurken geçmiş NBA müsabakalarının verileri ve takımların özet bilgilerinden yararlanılmıştır ve Markov model yaklaşımı ile basketbol müsabakaları ile tahminleme yapmanın tutarlı sonuçlar verebildiğine ulaşılmıştır (Štrumbelj ve Vračar (2012)).

Spor müsabakalarının tahminlenmesinde makine öğrenmesi kullanımı ile ilgili olarak;

Joseph ve arkadaşları Tottenham futbol kulübünün 1995-1997 yılları arasında oynadığı maçları incelemişlerdir. Çalışmada Bayesian Networks kullanılmıştır ve bazı alternatif makine öğrenmesine modellerine göre daha başarılı öngörüler elde etmişlerdir (Joseph vd. (2006)).

Regresyon ve nöral modellerin, spor sonuçlarını tahmin etmedeki doğruluğunu karşılaştırmaya yönelik bir araştırmada, 18 ± 0.5 yaşlarında bir grup 116 cirit atıcı incelenmiştir. Çalışmada Korelasyon matrisi ve regresyon analizinde; çapraz basamak, kolların ve gövdenin özgül gücü, abdominal kasların özgül gücü ve kavrama gücü kullanılarak doğrusal olmayan regresyon modelleri ve nöral modeller oluşturulmuştur. Sonrasında 2 yöntemle de Polonya Milli Takımı'ndan seçilen 20 cirit atıcı grubu için tahminleme yapılmış ve sonuçlar gerçek veriler ile karşılaştırılmıştır. Doğrusal olmayan regresyon modelleri ve 4-3-1 olarak yapılandırılan perceptron ağları ile tahminlenen değerler karşılaştırıldığında; mutlak hataların değeri arasındaki fark 12.68 m olarak tespit edilmiştir. Sonuç olarak, araştırma yapay sinir modeli ile ciddi oranda daha tutarlı tahminleme yapılabildiğini göstermiştir (Maszczyk vd. (2014)).

Konuyla ilgili ülkemizdeki çalışmalara örnek olarak da "Makine Öğrenmesi ile Spor Karşılaşmalarının Modellenmesi" çalışması gösterilebilir. Çalışmada genel olarak takımların ofansif ve defansif güçleri baz alınmış ve bu doğrultuda bulunulan sezonda attıkları ve yedikleri gollere dayalı basit bir model oluşturulmuştur (Karaoğlu (2015)).

Oyuncu performanslarının değerlendirilmesi ile ilgili bir çalışmada deep learning kullanılmıştır. Araştırmada spor haberleri kaynaklarından yapılan nitel taramalar ile oyun sonuçlarından çıkan nicel veriler kullanılarak derin öğrenmeye dayalı bir oyuncu değerlendirme modeli önerilmiştir. Yöntem olarak ise oyuncular belirli dönemlerde performanslarına göre pozitif veya negatif olarak sınıflandırılabileceği ve sonrasında aynı dönemdeki haber makaleleri kullanılır. Haber makalelerinin kullanılması aşamasında ise bir derin sinir ağı yapısı kullanılabileceği savunulmuştur. Sistem, Kore beyzbol ligine uygulanmış ve aradaki ilişki açıklanmıştır (Park vd. (2017)).

1.2 Tezin Amacı

Bu çalışma ile, spor sektöründeki bireysel (antrenör, menajer ve sporseverler) ve kurumsal (spor kulüpleri, oran hesaplayıcılar vb.) paydaşlar için, daha yüksek başarılı oranlarına sahip spor müsabakası tahminlemesi sağlamayı hedefliyoruz. Bu tahminleme sayesinde bireysel ve kurumsal kullanıcılar farklı kadro ve taktik seçimlerine göre, hava durumu, antrenör özellikleri, takımların mevcut ve geçmiş sezondaki performansları, hakem tarzı gibi birçok değişkeni dikkate alarak senaryo analizleri gerçekleştirebilecektir. Böylece, bu aracı kullanan takımların daha kazanma oranını ve alacakları puanlarda artış sağlanacaktır.

Tahminleme aracının spor sektöründe kurgulanmasının ardından, öncelikle finans, tarım ve enerji alanları üzerinde aynı aracın farklı versiyonlarının geliştirilmesi hedeflenebilir. Böylece aşağıdaki örneklerde piyasaya fayda sağlanması amaçlanacaktır:

Bankacılık sektöründe tarım alanına yönelik olarak 1.5 Milyar TL batık kredi bulunmaktadır. Bu durum, birçok bankanın kredi tahsis süreçlerindeki risk hesaplamalarında sektör bazlı ayrıma gitmemiş ve üretimden kaynaklı hesaplamaların risk belirleyiciler için detaylandırılmamış olmasından kaynaklıdır. 2014 yılında Malatya'da yaşanan don, ürünlerin %90'ının heba olmasına sebep olmuştur. 150-200 ton olarak beklenen kayısı üretimi, ancak 10 bin tona ulaşabilmiş ve Malatya ekonomisinin kaybı 360 Milyon Dolar olmuştur.

2015 yılında elektrik iletim hatlarında yaşanan problemden kaynaklı olarak 31 Mart tarihinde Türkiye genelinde 79 ili etkileyen bir elektrik kesintisi yaşanmıştır. Bu elektrik arızasının iletim hatlarında yaşanan yük hesaplama sorunundan meydana geldiği Enerji Bakanı tarafından açıklanmıştır. Bir günlük kesintinin ülke ekonomisine etkisi Ankara Sanayi Odası tarafından 700 Milyon Dolar olarak hesaplanmıştır. Bu örneklerde oluşan ekonomik zararın sebebi, verilerin doğru bir şekilde analiz edilmesi oluşabilecek farklı senaryoların fark edilmesi ve buna göre aksiyon alınması konusunda yaşanan aksaklıklardır. Gelecekte buna benzer kayıpları önlemek üzere, spor sektöründe edinilen algoritma tecrübesi farklı data içerikleri ile farklı sektörlere uygun geliştirme maliyetleri ile uygulanabilecektir.

1.3 Hipotez

Bu çalışmada mevcut veriler (kadro/oyuncu özellikleri, taktik yaklaşımlar, takım performansları, kart ve skor verileri, antrenör özellikleri, hakem etkisi, hava durumu etkisi, seyirci baskısı vb.) içinden en doğru tahmini yapmayı sağlayacak veriler ayıklanarak maç algoritmasına yansıtılacaktır. Projede, her bir müsabaka özelinde sonuca etki edecek verilerin korelasyonu ölçülmekte, geçmiş sonuçlar ile pozitif veya negatif korelasyonu olan veriler ile doğru olasılık hesabına ulaşmaya çalışılmaktadır. Bunun için Makine Öğrenmesi prensipleri uygulanmaktadır.

Çalışmanın öncelikle futbol olmak üzere, basketbol da dahil 20 farklı spor dalı için öncü olması amaçlanmaktadır.

Çalışmanın sahip olacağı teknik özellikler:

- Müsabaka bazında geçmiş verilere dair korelasyon analizi
- Makine Öğrenmesi prensiplerine dayanan yüksek başarı oranlı müsabaka sonucu olasılık analizi
- Farklı değişkenlerin etkisini değerlendirmeyi sağlayan senaryo analiz aracı

Çalışma ülke ve dünya için yenilik içeren bir ürün yeniliği ortaya koymaktadır.

2
MAKİNE ÖĞRENMESİ

Makine Öğrenmesinin Kullanım Alanları

Görüntü ve Konuşma Tanıma, Tıbbi Tanı, Tahmin, Sınıflandırma, Öğrenme Dernekleri, İstatistiksel Arbitraj, Ekstraksiyon, Regresyon

Bu projede, makine öğrenmesi tahmin ve istatistiksel arbitraja yönelik kullanılması amaçlanmaktadır. İlk çalışmada, geçmiş verilerden öğrenen makine öğrenmesi algoritmasıyla, gelecek futbol müsabakalarını tahminlemeye yönelik model geliştirilmesi üzerine çalışılacaktır.

2.1.1 Makine Öğrenmesinin Tahminleme için Kullanılması

Güncel tahmin en yeni makine öğrenme algoritmalarından biridir. Çeşitli amaçlar için tüm sektörlerde kullanılabilir.

Makine öğrenimi, sınıflandırma ve tahmin alanlarında umut verici sonuçlar veren akıllı metodolojilerden biridir. İyi öngörücü doğruluğu gerektiren genişleyen alanlardan biri, sektörün ciddi ekonomik boyutu nedeniyle spor tahminidir. Buna ek olarak, kulüp yöneticileri ve sahipleri sınıflandırma modelleri için uğraşırlar, böylece maçları kazanmak için gereken stratejileri anlayabilir ve formüle edebilirler. Bu modeller, oyunlarda yer alan, geçmiş maçların sonuçları, oyuncu performans göstergeleri ve muhalefet bilgileri gibi çeşitli faktörlere dayanmaktadır. Rory P.Bunker ve Fadi Thabtah, Yapay Sinir Ağı (YSA) uygulamasının spor sonuçları tahminine uygulanmasına odaklanmıştır. (Bunker ve Thabtah (2017) A Machine Learning framework for sport result prediction, Computing and Informatics.

İstatistiksel Arbitraj

Finansmanda, istatistiksel arbitraj, kısa dönemde tipik olan ve çok sayıda menkul kıymet içeren otomatik işlem stratejileri anlamına gelir. Bu tür stratejilerde, kullanıcı, tarihsel korelasyonlar ve genel ekonomik değişkenler gibi nicelikler temelinde bir dizi menkul kıymet için bir ticaret algoritması uygulamaya çalışır. Bu

ölçümler bir sınıflandırma veya tahmin problemi olarak kullanılabilir. Temel varsayım, fiyatların tarihsel ortalamaya doğru hareket edeceğidir. Makine Öğrenmesi ile gelecekteki bir zaman ufku boyunca beklenen getiri tahmin edilmeye çalışılabilir. Bu durumda, genellikle bir alım satım kararı almak için beklenen getiri tahminlerini kullanmak gerekir.

Christopher Krauss, Xuan Anh Do ve Nicolas Huck, S & P 500'deki tüm stokların gecikmeli getirisi konusunda, deep neural networks (DNN), gradient-boosted-trees (GBT) ve random forests (RAF) kullanarak, 1992'den 2015'e kadar, bir günlük hisse senedi piyasasının genel piyasadan daha iyi performans göstermesi olasılığına dayanarak günlük bir gün öncesindeki alım satım sinyalleri oluşturmuştur. Çalışmada, günlük %0.45'i aşan getiriye ulaşan sonuçlar elde etmişlerdir. (Krauss, Do ve Huck (2017) Deep neural networks, gradient-boosted trees, random forests: Statistical arbitrage on the S&P 500, European Journal of Operational Research, Volume 259, Issue 2, 1, Pages 689-702)

Öğrenme Yöntemine Göre Makine Öğrenmesi Algoritmaları

Öğrenme yöntemine göre makine öğrenmesi algoritmaları, Supervised learning (Gözetimli Öğrenme), Unsupervised Learning (Gözetimsiz Öğrenme), Semi-supervised Learning (Yarı Gözetimli Öğrenme) ve Reinforcement Learning (Takviye Öğrenme) olmak üzere 4 ana başlıkta sınıflandırılmaktadır.

1 - Supervised learning (Gözetimli Öğrenme)

- Nearest Neighbor
- Naive Bayes
- Decision Tree
- Linear Regression
- Support Vector Machines (SVM)
- Neural Networks

2 - Unsupervised Learning (Gözetimsiz Öğrenme)

- k-means clustering
- Association Rules

3 - Semi-supervised Learning (Yarı Gözetimli Öğrenme)

4 - Reinforcement Learning (Takviye Öğrenme)

- Q-Learning
- Temporal Difference (TD)
- Deep Adversarial Networks

Bu çalışmada supervised learning (gözetimli öğrenme) yöntemi kullanılacaktır. İlk çalışma olan skormetrede, belirlenen girdi değerlerinin geçmişte hangi output değerleri ile sonuçlandığı baz alınacak ve bu doğrultuda ilişkiler öğretilecektir.

2.1.2 Gözetimli Öğrenme

Bilinmeyen veriler üzerinde bilinen kalıpları keşfetmek isteyen bir makine öğrenme algoritması türüdür. Gözetimli öğrenme algoritmaları, hedef tahmin çıktısı ile girdi özellikleri arasındaki ilişkileri ve bağımlılıkları modellemeye çalışmaktadır, böylece önceki veri kümelerinden öğrendiği ilişkiler temelinde yeni veriler için çıktı değerlerini tahmin edebilmektedir.

Gözetimli öğrenmede, gözetimsiz öğrenmeden farklı olarak, geçmişteki sonuçlar bilinmektedir ve öğrenme işlemi girdiler ile bu sonuçların ilişkilendirilmesi ile gerçekleşmektedir. Örneğin, bir yıl içinde birinin kalp krizi geçirip geçirmeyeceğini tahmin etmek istediği durumda, önceki hastalarda, yaş, kilo, boy, kan basıncı gibi bir dizi veri bilinmektedir. Bu veriler dahilinde, kişilerin bir yıl içinde kalp krizi geçirip geçirmediği de bilinmektedir. Sonuç olarak, mevcut tüm veriler ile yeni bir kişinin bir yıl içinde kalp krizi geçirip geçirmeyeceğini öngörebilen bir model oluşturulmaktadır.

Gözetimli öğrenme sınıflandırma ve regresyon olmak üzere 2 kategoriye ayrılmaktadır.

Sınıflandırmada amaç, bir sonlu kümeler sınıfından bir sınıfa bir sınıf (veya etiket) atamaktır. Yani, yanıtlar kategorik değişkenlerdir. Uygulamalar arasında spam filtreleri, reklam önerme sistemleri ve görüntü ve konuşma tanıma bulunur. Bir hastanın bir yıl içinde kalp krizi geçirip geçirmeyeceğini tahmin etmek bir

sınıflandırma problemidir ve olası sınıflar doğru ve yanlıştır. Sınıflandırma algoritmaları genellikle nominal tepki değerleri için geçerlidir.

Regresyonda amaç, bir gözlem için sürekli bir ölçümü öngörmektir. Yani, yanıtlar değişkenleri gerçek sayılardır. Hisse fiyatlarını veya enerji tüketimlerini tahminlemek gibi uygulamalar içermektedir. Bu çalışma regresyon kategorisinde yer almaktadır.

Gözetimli öğrenme kullanılarak, çeşitli algoritmalar verimli bir şekilde eğitilebilir, modeller bir araya getirilebilir, model performanslarını değerlendirebilir, çapraz doğrulama yapabilir ve yeni veriler için yanıtlar tahmin edilebilir.

Çalışma Biçimlerine Göre Algoritmalar

En yaygın kullanılanlar incelendiği takdirde;

- Regression Algorithms (Regresyon Algoritmaları): Ordinary Least Squares Regression (OLSR), Linear Regression, Lojistik Regresyon, Stepwise Regression, Multivariate Adaptive, Regression Splines (MARS), Locally Estimated Scatterplot Smoothing (LOESS)
- Instance-based Algorithms (Örnek tabanlı Algoritmalar): k-Nearest Neighbor (kNN), Learning Vector Quantization (LVQ), Self-Organizing Map (SOM), Locally Weighted Learning (LWL)
- Regularization Algorithms (Düzenleme Algoritmaları): Ridge Regression, Least Absolute Shrinkage and Selection Operator (LASSO), Elastic Net, Least-Angle Regression (LARS)
- Decision Tree Algorithms (Karar Ağacı Algoritmaları): Classification and Regression Tree (CART), Iterative Dichotomiser 3 (ID3), C4.5 and C5.0 (different versions of a powerful approach), Chi-squared Automatic Interaction Detection (CHAID), Decision Stump, M5, Conditional Decision Tree
- Bayesian Algorithms (Bayes Algoritmaları): Naive Bayes, Gaussian Naive Bayes, Multinomial Naive Bayes, Averaged One-Dependence Estimators (AODE), Bayesian Belief Network (BBN), Bayesian Network (BN)

- Clustering Algorithms (Kümeleme Algoritmaları): k-Means, k-Medians, Expectation Maximisation (EM), Hierarchical Clustering
- Association Rule Learning Algorithms (İlişkilendirme Kuralı Öğrenme Algoritmaları): Apriori algorithm, Eclat algorithm
- Artificial Neural Network Algorithms (Yapay Sinir Ağı Algoritmaları): Perceptron, Back-Propagation, Hopfield Network, Radial Basis Function Network (RBFN)
- Deep Learning Algorithms (Derin Öğrenme Algoritmaları): Deep Boltzmann Machine (DBM), Deep Belief Networks (DBN), Convolutional Neural Network (CNN), Stacked Auto-Encoders
- Dimensionality Reduction Algorithms (Boyutluluk Azaltma Algoritmaları): Principal Component Analysis (PCA), Principal Component Regression (PCR), Partial Least Squares Regression (PLSR), Sammon Mapping, Multidimensional Scaling (MDS), Projection Pursuit, Linear Discriminant Analysis (LDA), Mixture Discriminant Analysis (MDA), Quadratic Discriminant Analysis (QDA), Flexible Discriminant Analysis (FDA)
- Ensemble Algorithms (Topluluk Algoritmaları): Boosting, Bootstrapped Aggregation (Bagging), AdaBoost, Stacked Generalization (blending), Gradient Boosting Machines (GBM), Gradient Boosted Regression Trees (GBRT), Random Forest

Bu çalışmada Artificial Neural Network Algorithms (Yapay Sinir Ağı Algoritmaları) kullanılacaktır. Yapay Sinir Ağları, biyolojik sinir ağlarının yapısı ve / veya işlevinden esinlenen modellerdir ve regresyon ve sınıflandırma problemleri için yaygın olarak kullanılan bir desen eşleştirme sınıfıdır.

3
YAPAY SİNİR AĞLARI

Ağırlıklı bağlantılar aracılığıyla birbirine bağlanan ve her biri kendi belleğine sahip işlem elemanlarından oluşan paralel ve dağıtılmış bilgi işleme yapılarıdır.

Gerçekte, yapay nöronlar genellikle, giriş özelliklerinin doğrusal bir kombinasyonu üzerinde doğrusal olmayan bir fonksiyon olarak uygulanır.

- Matematiksel olarak: $y \leftarrow f(W.x)$,
- W bir matristir
- x giriş özelliklerinin vektörüdür
- f, tanh gibi doğrusal olmayan bir işlevdir.

Yapay Sinir Ağları kullanılarak denetlenen ve denetlenmeyen algoritmalar vardır, en yaygın olarak bilinen, nöronların bağlanmış ve yönlendirilmiş bir grafiği olan ileri beslemeli sinir ağıdır.

Yapay Sinir Ağları Kullanım Alanları

Kontrol ve sistem tanımlama, görüntü ve ses tanıma, tahmin ve kestirim, arıza analizi, tıp, haberleşme, trafik, üretim yönetimi şeklindedir.

PETTERSSON ve NYQUIST'in Futbol maçlarının sonuçlarını tahmin etmek için derin öğrenme kullandığı araştırmada, 63 farklı ülkeden ve 3 turnuvadan geçmiş futbol maçları incelenmiştir. Çalışmada, tekrarlamalı yapay sinir ağı yapısı kullanılmıştır ve başarılı sonuçlar elde edilmiştir. Ancak, araştırma sonucuna daha detaylı veriler kullanılarak daha tutarlı sonuçlar elde edilebileceği eklenmiştir (Petterson ve Nyquist (2017)).

Yapay Sinir Ağları Özellikleri

- Öğrenme sırasında giriş ve çıkış bilgileri verilir, kurallar belirlenir.
- Hesaplama toplu olarak öğrenmeden sonra yapılır.

- Yavaştır ve hız donanıma bağlıdır. Bazı modeller, ciddi bilgisayar gücü ve alan gerektirebilir.
- Kapsamlı bir veri setine ihtiyaç vardır.

Yapay Sinir Yapısı

3.1.1 Girdiler

Hücreye dışarıdan verilen bilgilerdir.

3.1.2 Ağırlıklar

Girdinin hücre üzerindeki etkisini ifade etmektedir. Ağırlıkların değerleri, önemiyle orantılı değildir. Daha verimli bir model için ağırlıkları iyi belirlemek gerekmektedir.

3.1.3 Toplama Fonksiyonu (Birleştirme Fonksiyonu)

Hücreye gelen net bilgiyi hesaplar. Bunun için değişik fonksiyonlar kullanılmaktadır. En yaygın fonksiyon ağırlıklı toplamdır. Burada her gelen bilgi kendi ağırlığı ile çarpılarak toplanır. Böylece ağa gelen net girdi bulunmuş olur. Bu çalışmada da ağırlıklı toplam kullanılmıştır.

3.1.4 Aktivasyon Fonksiyonu

Hücreye gelen net girdi işlenir ve hücrenin bu girdiye karşılık üreteceği çıktı belirlenir. Aktivasyon fonksiyonu olarak genelde doğrusal olmayan bir fonksiyon seçilir. Bu çalışmada üssel aktivasyon fonksiyonu kullanılmıştır.

3.1.5 Hücrenin Çıktısı

Aktivasyon fonksiyonu sonucu belirlenen çıktı değeridir. Çıktı dış dünyaya, başka bir hücreye veya hücrenin kendisine tekrar girdi olarak gönderilir

Yapay Sinir Ağları İle Geliştirilen Modeller

3.1.6 Tek Katmanlı Algılayıcılar

Bu modelde yalnızca girdi ve çıktı katmanları bulunmaktadır. Bütün girdi üniteleri çıktı ünitelerine bir ağırlık değeri ile bağlıdır.

3.1.7 Basit Algılayıcılar

Sınıflandırma problemlerinde kullanılır ve bir sinir hücresine gelen birden fazla girdiyi alarak bir çıktı üretilmesini sağlamaktadır. Çıktılar bir veya sıfırdan oluşan mantıksal değerlerdir. Çıktı hesaplanırken eşik değer fonksiyonu kullanılır ve 0 veya 1 değerleri atanır.

3.1.8 Çok Katmanlı Algılayıcılar

Tek katmanlı ve basit algılayıcılar doğrusal olmayan çözümler konusunda yetersiz kalmaktadır. Bu ancak çok katmanlı algılayıcılar (MLP) ile sağlanabilmektedir. Bu çalışmada da çok katmanlı algılayıcılar kullanılmıştır. Doğrusal olmayan aktivasyon fonksiyonuna sahip birçok nöronun birbirine bağlıdır. Genel olarak yapay sinir ağları ileri beslemeli ve geri beslemeli ağlar olarak ikiye ayrılmaktadır. Bu çalışmada ileri beslemeli ağlar kullanılmıştır.

4
FUTBOL MODELİ KURGULANMASI

4.1 Parametre Seçimi

Bu çalışmada da makine öğrenmesi yöntemleri kullanılarak futbol müsabakalarının sonuçlarının olasılık bazında tahminlenmesi amaçlanmıştır ve bu kapsamda kullanılacak parametreler için maç sonu ve maç içi istatistikleri kullanılarak 2 grup oluşturulmuştur.

4.1.1 Maç Sonu Parametreleri - Vol1

Geçmiş futbol müsabakalarının sonuçlarından yola çıkılarak elde edilen parametrelerdir. Tablo 4.1'de gösterilmiştir.

Tablo 4. 1 Maç sonu parametreleri

ID	Analiz Tipi	Parametre
1	Aralarındaki Maçlar	Ev sahibi takımın, evinde yapılan maçlarda attığı gol ort
2	Aralarındaki Maçlar	Deplasman takımın, ev sahibi takımın evinde yapılan maçlarda attığı gol ort
3	Aralarındaki Maçlar	Deplasman takımın, evinde yapılan maçlarda attığı gol ort
4	Aralarındaki Maçlar	Ev sahibi takımın, deplasman takımın evinde yapılan maçlarda attığı gol ort
5	Gol Analizi	Ev Sahibi Takımın Bu Sezon Evinde Attığı Gol Sayısı Ortalaması
6	Gol Analizi	Ev Sahibi Takımın Bu Sezon Evinde Yediği Gol Sayısı Ortalaması
7	Gol Analizi	Deplasman Takımının Bu Sezon Deplasmanda Attığı Gol Sayısı Ortalaması
8	Gol Analizi	Deplasman Takımının Bu Sezon Deplasmanda Yediği Gol Sayısı Ortalaması
9	Gol Analizi	Ev Sahibi Takımın 1 Önceki Sezon Evinde Attığı Gol Sayısı Ortalaması
10	Gol Analizi	Ev Sahibi Takımın 1 Önceki Sezon Evinde Yediği Gol Sayısı Ortalaması
11	Gol Analizi	Deplasman Takımının 1 Önceki Sezon Deplasmanda Attığı Gol Sayısı Ortalaması
12	Gol Analizi	Deplasman Takımının 1 Önceki Sezon Deplasmanda Yediği Gol Sayısı Ortalaması
13	Gol Analizi	Ligde Bu Sezon Atılan Gol Sayısının Ortalaması
14	Lig Performansı – Son X	Ev Sahibi Takımın Son 5 Maçtaki Puan Oranı
15	Lig Performansı – Son X	DeplasmanTakımın Son 5 Maçtaki Puan Oranı
16	Lig Performansı	Ev Sahibi Takımın Maç Başına Puan Ortalaması
17	Lig Performansı	Deplasman Takımının Maç Başına Puan Ortalaması
18	Lig Performansı	Ev Sahibi Takımın Ligdeki Sıralamasının Normalize Değeri
19	Lig Performansı	Deplasman Takımının Ligdeki Sıralaması Normalize Değeri
20	Bahis Oranları	İddaa Maç Sonucu 1 Oranı
21	Bahis Oranları	İddaa Maç Sonucu 0 Oranı
22	Bahis Oranları	İddaa Maç Sonucu 2 Oranı
23	Skor	Ev Sahibi Takım Maç Sonucu Skoru
24	Skor	Deplasman Takımı Maç Sonucu Skoru

4.1.2 Maç İçi Parametreler - Vol2

Maç içi istatistiklerden yola çıkılarak elde edilen parametrelerdir. Tablo 4.2'de gösterilmiştir.

Tablo 4. 2 Maç içi parametreler

ID	Analiz Tipi	Parametre
1	Kadro	Ev Sahibi Kulüp Değeri
2	Kadro	Deplasman Kulüp Değeri
3	Kadro	Ev Sahibi Takım Boy Ortalaması
4	Kadro	Deplasman Takımı Boy Ortalaması
5	Kadro	Ev Sahibi Takım Yaş Ortalaması
6	Kadro	Deplasman Takımı Yaş Ortalaması
7	Kadro	Ev Sahibi Takım İlk 11 Takım Değerleri
8	Kadro	Deplasman Takımı İlk 11 Takım Değerleri
9	Maç İçi İstatistikler	Ev Sahibi Takımın Ortalama Topa Sahip Olma Oranı
10	Maç İçi İstatistikler	Deplasman Takımının Ortalama Topa Sahip Olma Oranı
11	Maç İçi İstatistikler	Ev Sahibi Takımın Başarılı Pas Oranı
12	Maç İçi İstatistikler	Deplasman Takımının Başarılı Pas Oranı
13	Maç İçi İstatistikler	Ev Sahibi Takımın Maç Başı Korner Oranı
14	Maç İçi İstatistikler	Deplasman Takımının Maç Başı Korner Oranı
15	Maç İçi İstatistikler	Ev Sahibi Takımın Şut Sayısı Oranı
16	Maç İçi İstatistikler	Deplasman Takımının Şut Sayısı Oranı
17	Maç İçi İstatistikler	Ev Sahibi Takımın Kaleyi Bulan Şut Sayısı Oranı
18	Maç İçi İstatistikler	Deplasman Takımının Kaleyi Bulan Şut Sayısı Oranı
19	Maç İçi İstatistikler	Ev Sahibi Takımın Kırmızı Kart Sayısı
20	Maç İçi İstatistikler	Deplasman Takımın Kırmızı Kart Sayısı
21	Maç İçi İstatistikler	Ev Sahibi Takımın İlk Yarı Kazanma Oranı
22	Maç İçi İstatistikler	Deplasman Takımının İlk Yarı Kazanma Oranı
23	Maç İçi İstatistikler	Ev Sahibi Takımın Hava Topu Başarı Oranı

4.1.3 Skorlama

Regresyon veya gruplandırma amacıyla işaretlenmiş sonuçların kullanılan parametreler ile olan ilişkisinin skorlanması işlemidir.

Skorlama yöntemleri

Bilgi Kazanımı (Information Gain)

Beklenen bilginin miktarı (entropinin azalması).

Bilgi Kazanımı Oran (Information Gain Ratio)

Elde edilen ilişkisiel bilgi ile bilginin kendisi arasındaki oran. Karar Verme Ağacı üzerinde karar verme becerisini arttırabilmek olabildiğince ayrı değerlerin belirlenmesine yönelik bir değerdir.

Gini

Frekans dağılımında farklı değerleri arasındaki eşitsizliği (inequality) gösterir

Anova

Bir parametrenin farklı sonuç (Sınıf) içindeki ortalama değerlerinin farkıdır.

Chi2

Parametre ve sonuç sınıfı arasındaki ilişkinin chi-square (Ki-Kare) cinsinden değeridir.

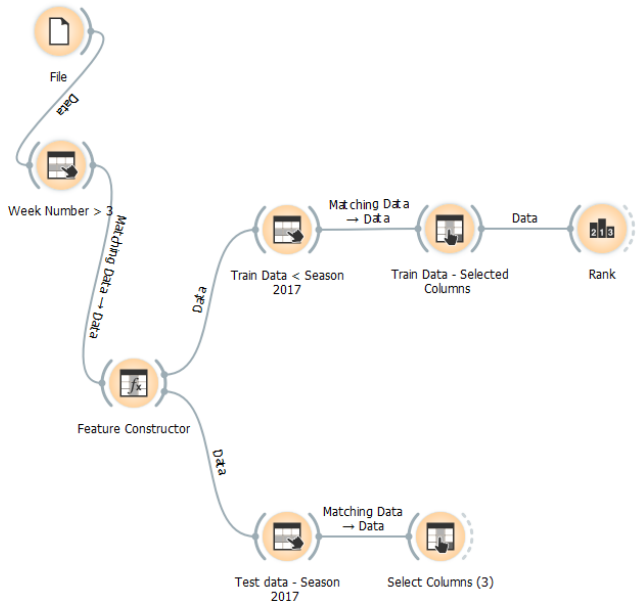
ReliefF

Bir parametrenin benzer veri örneklerindeki farkı gösterebilme gücü.

Fcbf

Entropi temelli, parametreler arasındaki korelasyonun yarattığı fazlalığı gösterir.

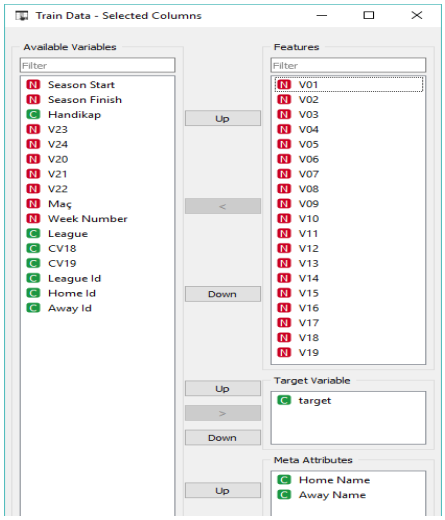
-Yöntem:



Şekil 4. 1 İlk model kurgulama yöntemi

Şekil 4.1'de gösterildiği gibi, dosya Orange aracına aktarılmış, 3. haftadan sonraki veriler için parametreler üretilmiş, eğitim ve test verileri ayrıştırılmıştır.

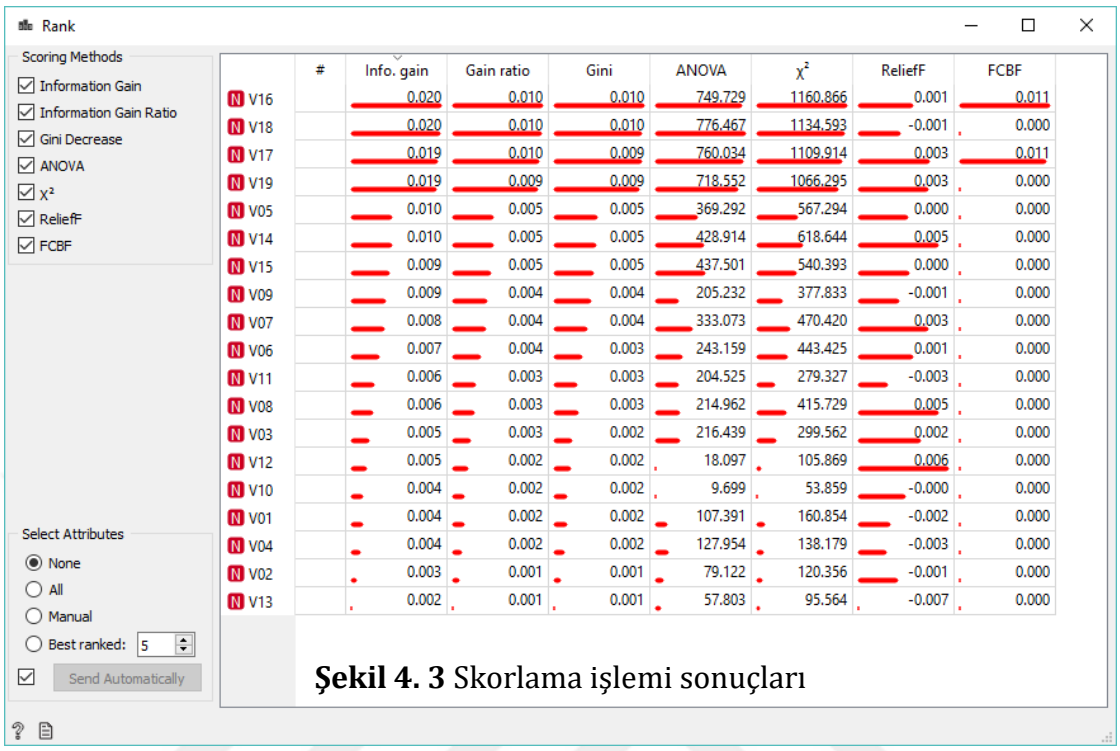
Şekil 4.2'de gösterildiği gibi parametreler seçilmiştir:



Şekil 4. 2 Python ile skorlama yöntemi

Skorlama İşlemi

İşlem sonrası Şekil 4.3'deki sonuçlar elde edilmiştir:



	#	Info. gain	Gain ratio	Gini	ANOVA	χ^2	ReliefF	FCBF
V16		0.020	0.010	0.010	749.729	1160.866	0.001	0.011
V18		0.020	0.010	0.010	776.467	1134.593	-0.001	0.000
V17		0.019	0.010	0.009	760.034	1109.914	0.003	0.011
V19		0.019	0.009	0.009	718.552	1066.295	0.003	0.000
V05		0.010	0.005	0.005	369.292	567.294	0.000	0.000
V14		0.010	0.005	0.005	428.914	618.644	0.005	0.000
V15		0.009	0.005	0.005	437.501	540.393	0.000	0.000
V09		0.009	0.004	0.004	205.232	377.833	-0.001	0.000
V07		0.008	0.004	0.004	333.073	470.420	0.003	0.000
V06		0.007	0.004	0.003	243.159	443.425	0.001	0.000
V11		0.006	0.003	0.003	204.525	279.327	-0.003	0.000
V08		0.006	0.003	0.003	214.962	415.729	0.005	0.000
V03		0.005	0.003	0.002	216.439	299.562	0.002	0.000
V12		0.005	0.002	0.002	18.097	105.869	0.006	0.000
V10		0.004	0.002	0.002	9.699	53.859	-0.000	0.000
V01		0.004	0.002	0.002	107.391	160.854	-0.002	0.000
V04		0.004	0.002	0.002	127.954	138.179	-0.003	0.000
V02		0.003	0.001	0.001	79.122	120.356	-0.001	0.000
V13		0.002	0.001	0.001	57.803	95.564	-0.007	0.000

Şekil 4. 3 Skorlama işlemi sonuçları

Sonuç

Sonuçlardan

V16: Ev Sahibi Takımın Maç Başına Puan Ortalaması

V17: Deplasman Takımının Maç Başına Puan Ortalaması

V18: Ev Sahibi Takımın Ligdeki Sıralamasının Normalize Değeri

V19: Deplasman Takımının Ligdeki Sıralaması Normalize Değeri

değerlerinin sonuçla olan ilişkisinin diğer parametrelere göre çok yüksek olduğu gözükmektedir.

Ayrıca takip eden;

V05: Ev Sahibi Takımın Bu Sezon Evinde Attığı Gol Sayısı Ortalaması

V14: Ev Sahibi Takımın Son 5 Maçtaki Puan Oranı

V15: DeplasmanTakımın Son 5 Maçtaki Puan Oranı

V09: Ev Sahibi Takımın 1 Önceki Sezon Evinde Attığı Gol Sayısı Ortalaması

parametrelerinin de sonuç üzerinde kısmı bir etkisi olduğu gözlenmiştir.

Bu parametrelerin daha etkin hale getirilmesi gerekmektedir. Korelasyon analizine geçilmelidir.

4.1.4 Korelasyon Analizi

Sonraki aşamada bağımlı/tekrar eden parametrelerin tespit edilmesi için korelasyon analizi uygulanmıştır. Pearson/Spearman tarafından geliştirilen parametre çiftleri arası monolitik ilişkiyi gösteren testler uygulanmıştır.

Pearson Korelasyon

Şekil 4. 4 Pearson korelasyon sonuçları

%70 üzeri korelasyonlar sırasıyla;

- V15 (DeplasmanTakımın Son 5 Maçtaki Puan Oranı) – V17 (Deplasman Takımının Maç Başına Puan Ortalaması)

- V09 (Ev Sahibi Takımın 1 Önceki Sezon Evinde Attığı Gol Sayısı Ortalaması) – V10 (Ev Sahibi Takımın 1 Önceki Sezon Evinde Yediği Gol Sayısı Ortalaması)
- V14 (Ev Sahibi Takımın Son 5 Maçtaki Puan Oranı) – V16 (Ev Sahibi Takımın Maç Başına Puan Ortalaması)

şeklinde tespit edilmiştir.

Sperman Korelasyon

Şekil 4. 5 Sperman korelasyon sonuçları

Sonuç

Sonuçlardan parametre çiftleri arasında korelasyon oranları yüksek gözlenmiştir.

Bu seçilen parametrelerin birbirinden bağımsız olmadıkları ve regresyon/gruplandırma modelinde tekrarlara yol açtığı anlaşılmıştır. Bağımsız parametre setlerinin kurgulanması gerekmektedir.

4.2 Veri Setinin Model Kurgusundan Önce İşlenmesi

Son 10 yıllık geçmiş maç sonuçları incelenmiştir.

02.03.2007 - 11.03.2020 tarihleri arasındaki 118.716 maç incelendiğinde,

- Ev sahibi takımın veya deplasman takımının 7'den fazla gol attığı maç sayısı: **73 (0,0006)**
- Ev sahibi takımın veya deplasman takımının 6'dan fazla gol attığı maç sayısı: **255 (0,002)**
- Ev sahibi takımın veya deplasman takımının 5'ten fazla gol attığı maç sayısı: **1.098 (0,009)**
- Ev sahibi takımın veya deplasman takımının 4'ten fazla gol attığı maç sayısı: **3.819 (0,032)**
- Ev sahibi takımın veya deplasman takımının 3'ten fazla gol attığı maç sayısı: **12.356 (0,10)**
- Ev sahibi takımın veya deplasman takımının 2'den fazla gol attığı maç sayısı: **34.014 (0,28)**
- Ev sahibi takımın veya deplasman takımının 1'den fazla gol attığı maç sayısı: **72.995 (0,61)**
- Ev sahibi takımın veya deplasman takımının 0'dan fazla gol attığı maç sayısı: **10.919 (0,92)**
- Ev sahibi takımın ve deplasman takımının 0 gol attığı maç sayısı: **9.797 (0,08)**

Bu kapsamda 7'den fazla gol atılan maçlar göz ardı edilebilir. Ev sahibi takım gol sayısı ve deplasman takımı gol sayılarını minimum 0, maksimum 7 gol olacak şekilde sınırlandırabiliriz. Ev sahibi takımın evinde yapılan maçlarda attığı gol ort V1, V2, V3, V4, V5, V6, V7, V8, V9, V10, V11, V12, V13 parametreleri gol bazlı parametreler olduğu için, öğrenme modeline minimum 0 ve maksimum 7 olan değerleri almak, 7 üzeri değerleri 7'ye sabitlemek uygun görülmüştür.

Puan bazlı parametrelerde ise minimum 0, maksimum 3 sınırları uygun görülmüştür. Bu kapsamda V14, V15, V16 ve V17 parametreleri için minimum değer 0 maksimum değer 3 olarak belirlenmiştir.

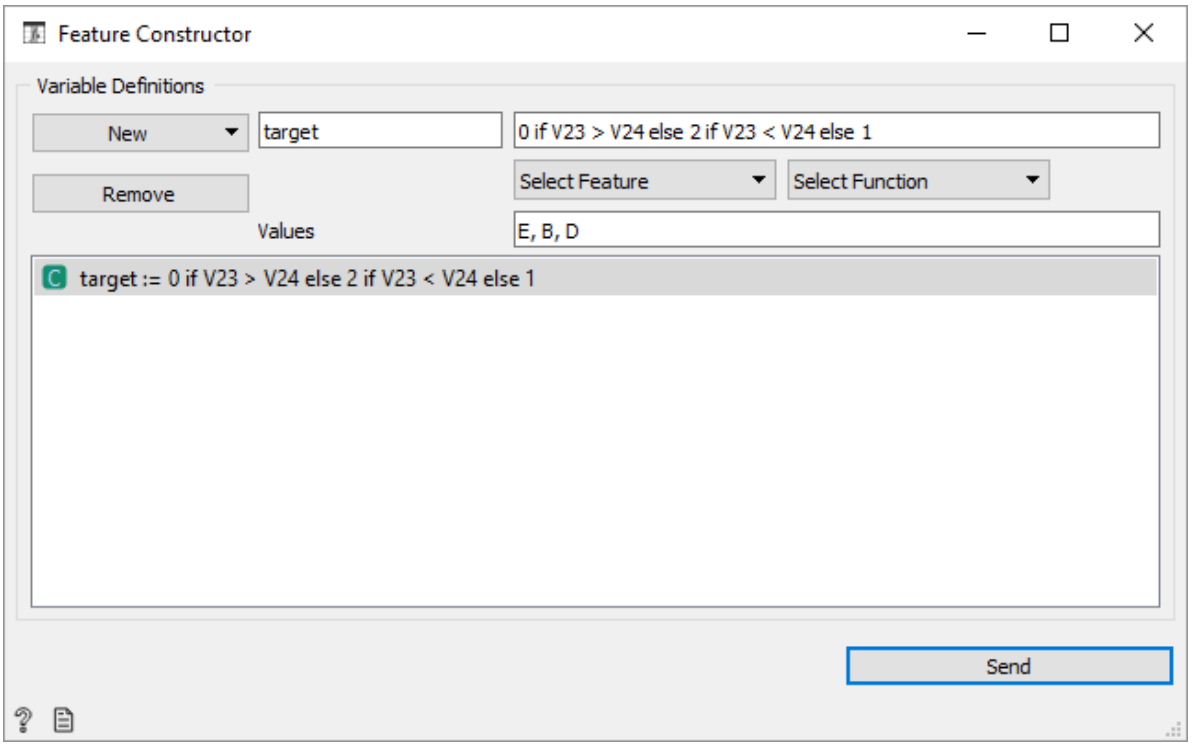
Sıralama bazlı parametrelerde ise, parametrenin öğrenme modelinde anlamlı hale gelebilmesi için ilk sıraya 0, son sıraya 1 olarak değer verip, aradaki sıraları 0 ve 1'e göre olan uzaklıklarına göre belirlemek uygun görülmüştür. Bu kapsamda V18 ve V19 parametreleri için minimum değer 0, maksimum değer 1 olarak belirlenmiştir.

4.3 Hedef Belirlenmesi

Maç sonucunu içeren V23 ve V24 kolonları üzerinden hedef sınıf Şekil 4.6'daki gibi oluşturulmuştur.

V23: Ev Sahibi Takım Maç Sonucu Skoru

V24: Deplasman Takımı Maç Sonucu Skoru



Şekil 4. 6 Python hedef seçimi

4.4 Eğitim Ve Test Verisi Belirlenmesi

Eğitim Verisi (Training Data): Algoritmanın öğrenmesi için sunulan gözlemler dizisidir. Algoritma ile bu veri baz alınarak çıkarımlar yapılır.

Çoğu öğrenme sistemi genellikle öğrenme için kullanılan eğitim setlerinin dengeli olduğunu varsayar. Ancak eğitim ve test verisi arasında uç dağılımlar da olabilir, belli özellikte veriler eğitim setinde toplanabilir. Bu da yanlış öğrenmeye neden olur. Bu nedenle veri setini geniş tutmak gerekir. (Gustavo vd. (2004))

Test Verisi (Test Data): Modelin başarısını test etmek için kullanılan veri setidir. Eğitim esnasında saklanır, eğitim bittikten sonra etiketsiz olarak algoritmaya verilerek algoritmanın etiketlenmeyen veriler hakkında tahminlerde bulunması beklenir.

Overfitting (Aşırı Uyum)

Dietterich'e göre öğrenme algoritmasındaki temel problem, gürültüden kaynaklı olarak, öğrenme modelinin veri setindeki gerçek ilişkiyi bulmak yerine tekrar eden verilerin çeşitli özelliklerini ezberleyerek overfitting yapmasıdır. (Dietterich (1995))

Gözetimli makine öğrenmesinde, veri setinin sınırlı ve tekrarlı olması ve öğrenmenin buna dayalı ilişki kuracak şekilde kurgulanması sonucunda gerçekleşir. Buna verisetine gerekenden daha fazla parametre uygulanması neden olabilir.

Overfitting durumunda, algoritma çalıştığımız veri seti üzerinde harika sonuçlar verir ancak hiç görmediği yeni veri setleri üzerinde başarısız tahminler yapar.

Yani eğitim hatası düşüktür, ancak test hatası yüksektir.

Bunun sebelerinden birisi noise (gürültü)'dür. Veri setindeki gerçek ilişkilerdense gürültü öğretilir.

Underfitting (Yetersiz Uyum)

Algoritmanın bazı ilişkileri hiç öğrenmediği, baştan atladığı yani öğrenmeye dahil edilmeyen verilerin olduğu koşullarda da underfitting durumu gerçekleşir. Bu durumda model eğitim (training) veri seti üzerinde çok basit olarak kurgulanmıştır. Bu nedenle model daha önce hiç karşılaşmadığı veri seti üzerinde başarısız sonuçlar üretir. Tahmin edilen değer ile gerçek değer arasında büyük farklar oluşur.

Makine öğrenmesi modellerinin sistematiği genelde şöyledir: Veri eğitim datası ve test verisi şeklinde ayrıştırılır. Makine öğrenmesi modelleri verinin bir kısmıyla eğitilir, bir kısmıyla da daha önce karşılaşmadığı test verisi üzerinde tahminler yaptırılır.

Jabbar ve Khan'a göre underfitting, overfitting'in tam tersidir ve modelin tekrarlayan veri setlerini ezberlemesinden ziyade ilişki bulamaması durumunda meydana gelir. Bu durumda başarı oranı da düşük olacaktır. Jabbar ve Khan bu

duruma karşı penaltı yöntemleri ve eğitimin erken durdurulması yöntemlerini önermiştir. (Jabbar ve Khan (2015))

Validasyon

Eğitim verisinin bir kısmına (%20'sine) hiç bakılmaz ve bu kısım validation data olarak adlandırılır, bu data üzerinden tahminleme yapıp modelin denenmesine validasyon metodu denir.

Bu metodun bazı sakıncaları bulunmaktadır. Bütün veri öğrenme sürecine dahil edilmez bir bölümü validasyon için ayrılır, bu durumda eksik veri underfitting problemine sebep olabilir.

Eğitimden geçen veri iyi bir örneklem olmayabilir.

Cross validation (Çapraz Doğrulama)

Validasyon aşamasındaki olası underfitting sorununu önleyebilmek için çapraz doğrulama yöntemi kullanılabilir. Literatürde çeşitli çapraz doğrulama metotları var.

Moore ve Lee'ye göre, model seçimi, makine öğrenmesinin en önemli aşamalarından biridir. Bir veri seti ve bu veri seti üzerinden tahminleme yapan bir dizi model arasından en başarılı precision ve recall değerlerini veren modeli seçmek gerekmektedir. Bu aşamada da Cross validation (Çapraz Doğrulama) başarılı bir yöntemdir. (Moore ve Lee (2015))

K-fold Yöntemi

1. deney için veri setinin yüzde 20'lik kısmını test (validation) için yüzde 80'lik kısmını eğitim için ayrılır.

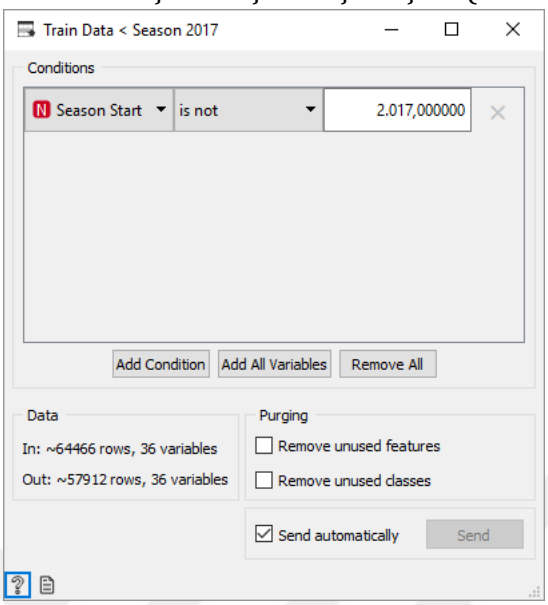
2. deney için veri setinin farklı bir yüzde 20'lik kısmı test için yüzde 80'lik kısmı eğitim için ayrılır ve bu prosedür 3. 4. 5. deneyler için de gerçekleştirilir. Günün sonunda modellerin başarı oranlarının ortalaması alınır.

Sonuç

Modelimizin karmaşıklığı arttıkça overfitting (aşırı uyum) artar underfitting (yetersiz uyum) azalır.

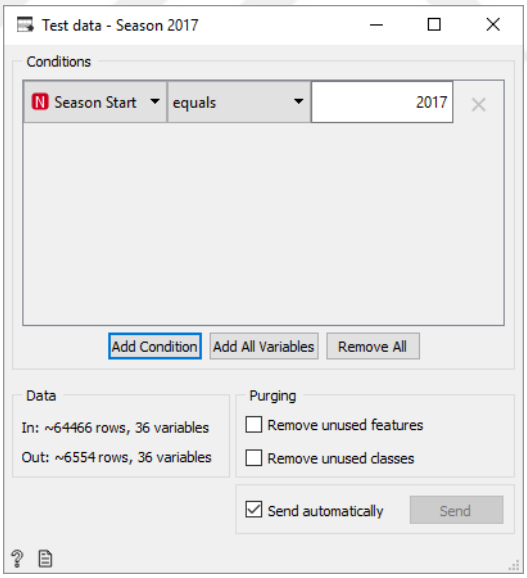
2008 – 2017 yılları arasındaki futbol müsabakalarının eğitim verisi olarak, 2017 – 2018 müsabakalarının ise Test verisi olarak kullanılmasına karar verilmiştir.

Eğitim verisi olarak 2017'den küçük maçlar seçilmiştir. (57912 kayıt)



Şekil 4. 7 Python eğitim verisi seçimi

Test verisi olarak 2017 yılındaki maçlar seçilmiştir. (6554 kayıt)

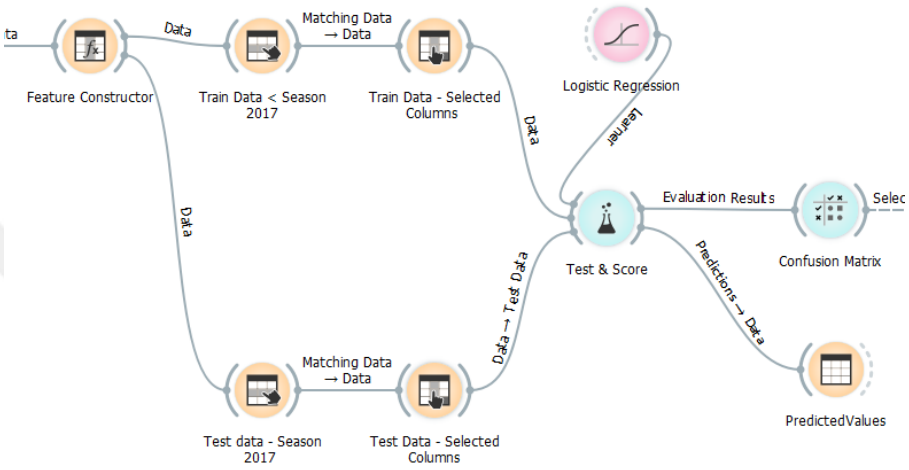


Şekil 4. 8 Python test verisi seçimi

5
TAHMİNLEME MODELİ

5.1 Lojistik Regresyon

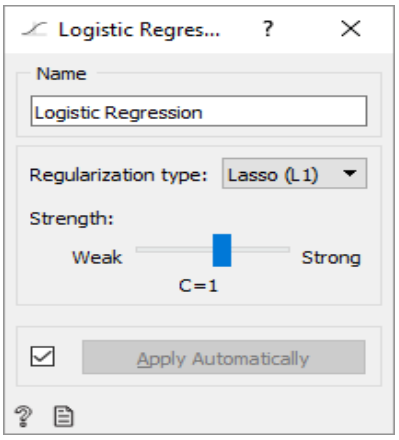
Şekil 5.1'de gösterildiği gibi, parametreler üretilmiş, eğitim ve test verileri ayrıştırılmış, lojistik regresyon için test sonuçları ve hata matrisi oluşturulmuştur.



Şekil 5. 1 Python tahminleme modeli lr yöntemi

Hiperparametreler:

Learner hiperparametreleri Şekil 5.2'deki gibi seçilmiştir. Bu parametrelerin farklı kombinasyonların sonuçlarda önemli bir değişiklik gözlenmemiştir.



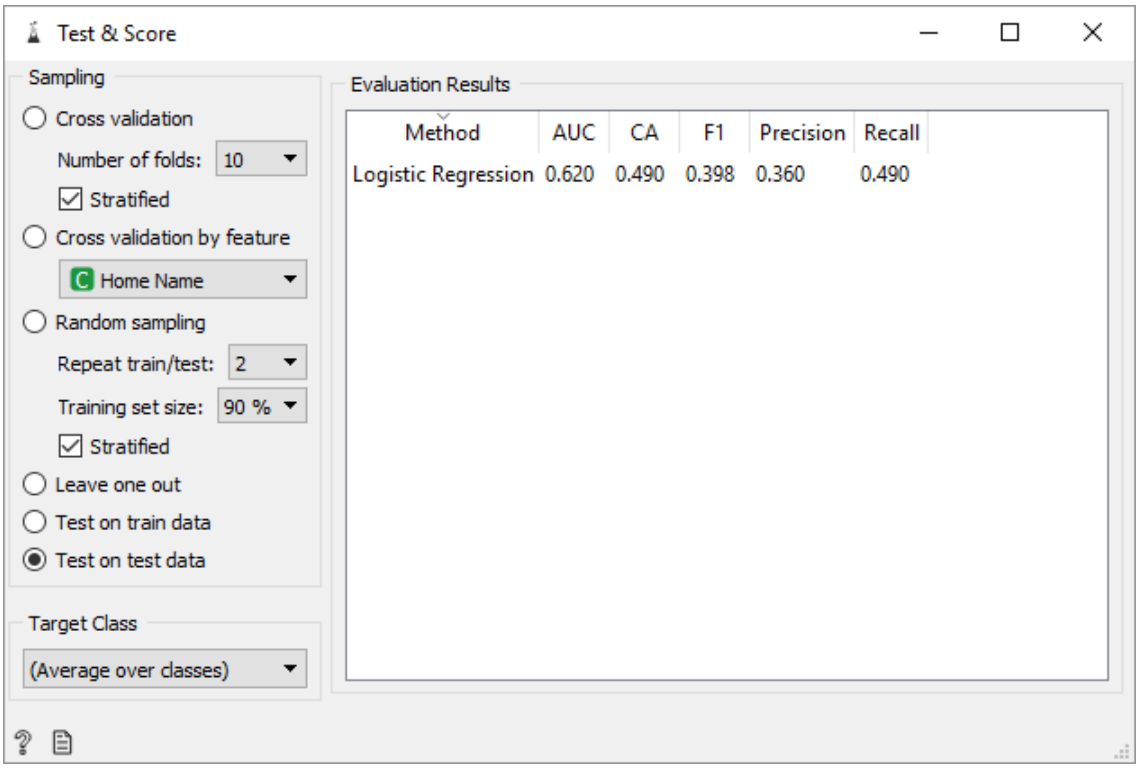
Şekil 5. 2 Python tahminleme modeli lr hiperparametre

5.2 Lojistik Regresyon Sonuç & Validasyon

Öğrenme ve tahminleme işlemi sonrası;

- 2017 öncesi sezonlar için öğrenme uygulanıp
- 2017 sezonundaki maçlar için test edildiğinde

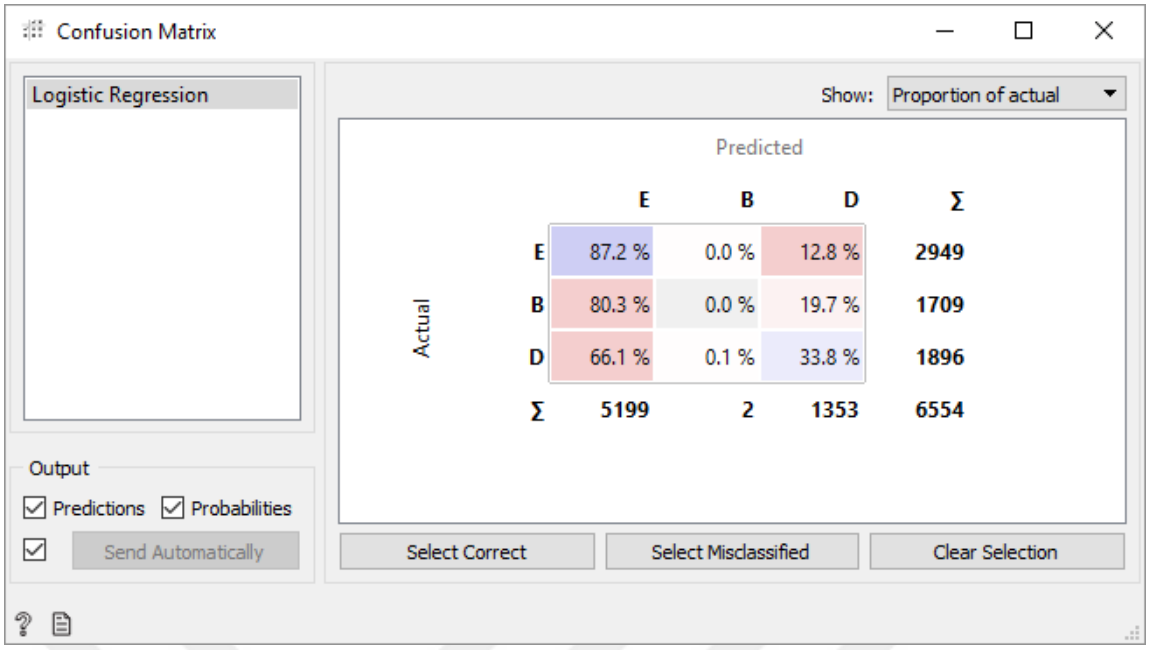
Şekil 5.3'deki sonuçlar ile karşılaşılmıştır



Şekil 5. 3 Tahminleme modeli lr sonuçları

- **Recall:** 0,49
  - Gerçekleşen sonucu modelin doğru bilme oranı **%49** olarak tespit edilmiştir.
- **Precision:** %36
  - Yani model tahminlerinin gerçeklemesi doğruluk oranı **%36** seviyesindedir.

Hata Matrisi



Şekil 5. 4 Tahminleme modeli lr hata matrisi

Sınıf değerleri bazındaki doğruluk değerleri Şekil 5.4'deki gibi bulunmuştur.

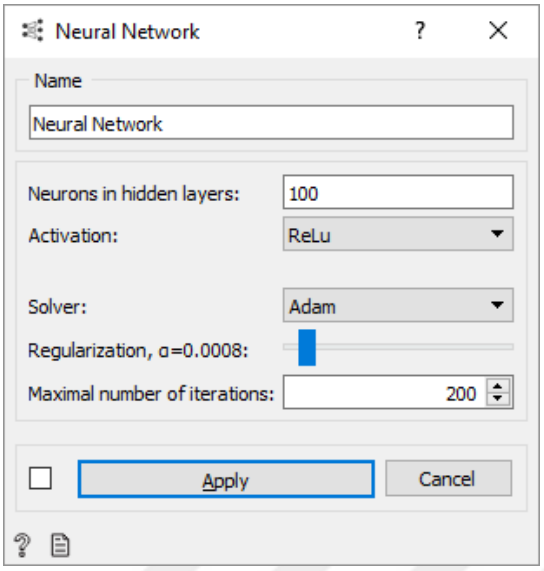
Test verisinde,

- Ev sahibinin kazandığı maçlar **% 87,2**
- Beraberlikle biten **% 0**
- Deplasman takımının kazandığı maçlar ise **% 33,8** oranında tahminlenebilmiştir.

5.3 Neural Network (Sinir Ağları)

Learner hiperparametreleri Şekil 5.5'deki gibi seçilmiştir. Tek katmanlı yapıda bu parametrelerin farklı kombinasyonlarının sonuçlarda önemli bir değişikliği gözlenmemiştir.

Hiperparametreler:



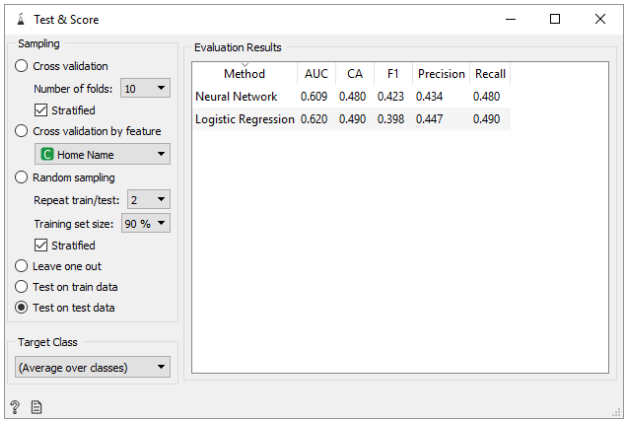
Şekil 5. 5 Tahminleme modeli ann hiperparametre seçimi

5.4 Neural Network (Sinir Ağları) Sonuç & Validasyon

Öğrenme ve tahminleme işlemi sonrası;

- 2017 öncesi sezonlar için öğrenme uygulanıp
- 2017 sezonundaki maçlar için test edildiğinde

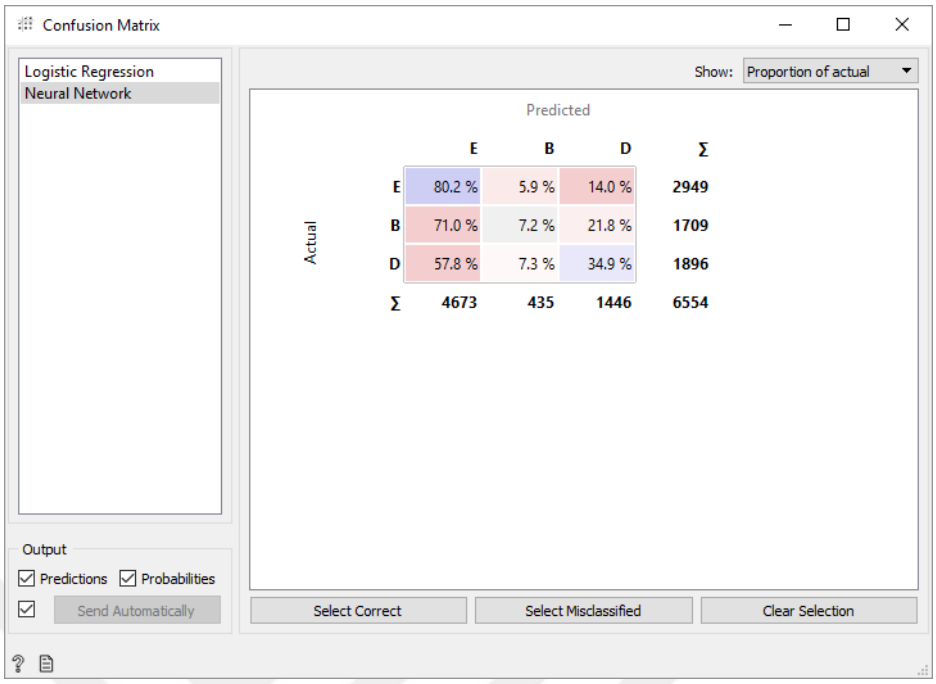
Şekil 5.6'daki sonuçlar ile karşılaşılmıştır.



Şekil 5. 6 Tahminleme modeli ann sonuçları

- **Recall:** 0,48
  - Gerçekleşen sonucu modelin doğru bilme oranı **%48** olarak tespit edilmiştir.
- **Precision:** %39.8
  - Yani model tahminlerinin gerçeklemesi doğruluk oranı **%39.8** seviyesindedir.

Hata Matrisi



Şekil 5. 7 Tahminleme modeli ann hata matrisi

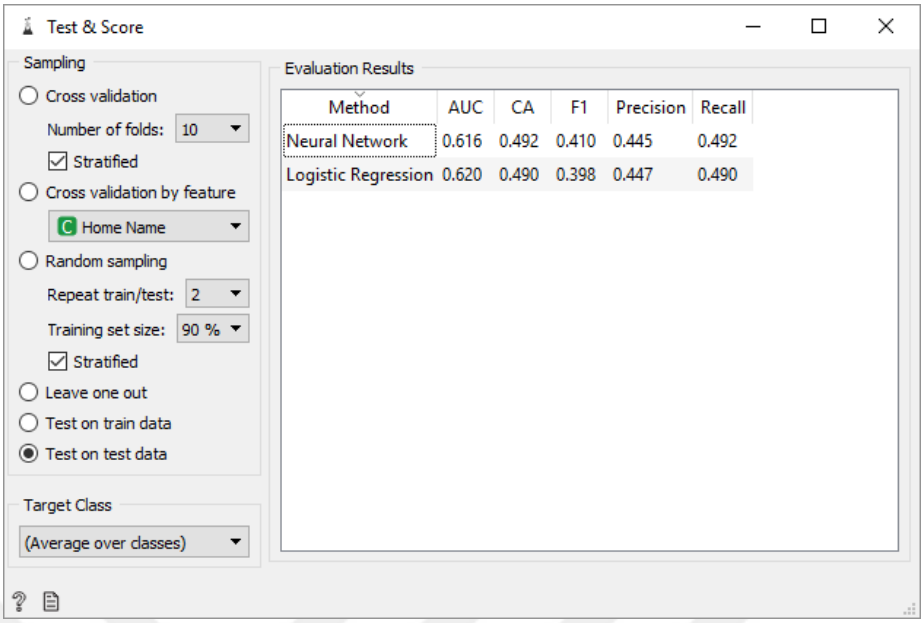
Test verisinde,

- Ev sahibinin kazandığı maçlar **% 80,2**
- Beraberlikle biten **% 7,2,**
- Deplasman takımının kazandığı maçlar ise **% 34,9** oranında tahminlenebilmiştir. Katman sayısı 2'e çekilerek deney tekrarlanmıştır.

Hiperparametreler:

Katman sayısı 2'ye çıkarılıp, Nöron sayıları 64'e düşürülüp tekrar denenmiştir.

Öğrenme ve tahminleme işlemi sonrası Şekil 5.8'deki sonuçlar ile karşılaşılmıştır:



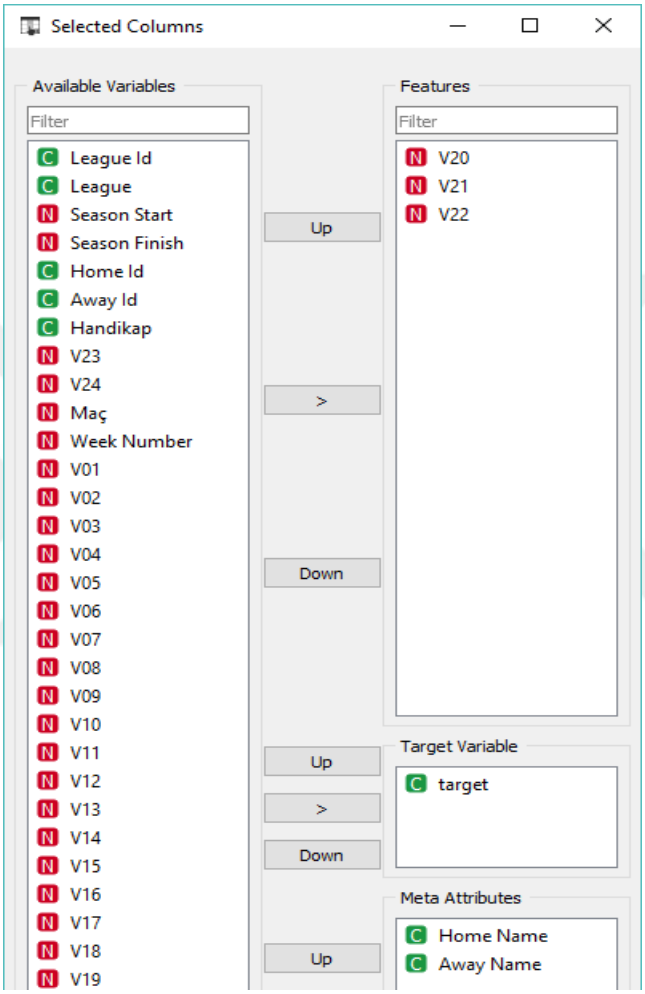
Şekil 5. 8 Tahminleme modeli ann sonuçları - 2

Deney bu parametreler ile tekrar yapıldığında;

- **Recall:** % 49.2
  - Gerçekleşen sonucu modelin doğru bilme oranı **%48** olarak tespit edilmiştir.
- **Precision:** %44.5
  - Yani model tahminlerinin gerçeklemesi doğruluk oranı **%39.8** seviyesindedir.

6
İDDAA MODELİ

Parametre olarak İddaa oranları seçilmiştir Şekil 6.1'deki V20 İddaa Maç Sonucu 1 oranını, V21 İddaa Maç Sonucu 0 oranını, V22 İddaa Maç Sonucu 2 oranını ifade etmektedir.

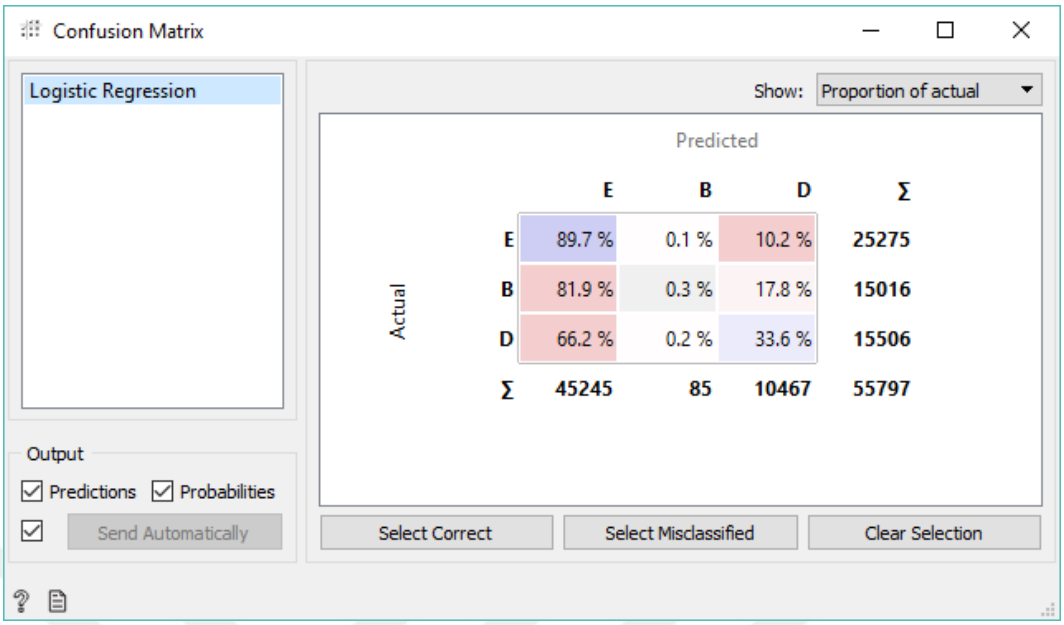


Şekil 6. 1 İddaa modeli parametre seçimi

Şekilde görüldüğü gibi V20, V21 ve V22 parametreleri kullanılmıştır. V20: İddaa Maç Sonucu 1 Oranı, V21: İddaa Maç Sonucu 0 Oranı, V22 İddaa Maç Sonucu 2 oranıdır.

6.1 Lojistik Regresyon & Başarı Oranları

Recall - 0,501 olarak hesaplanmıştır. Yani aranan sonuç için ortalama doğruluk %50 seviyesindedir.



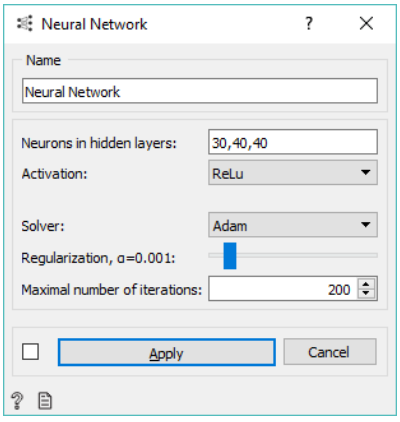
Şekil 6. 2 İddaa modeli lr hata matrisi

Sınıf değerleri bazındaki doğruluk değerleri Şekil 6.2'deki gibi bulunmuştur.

- Ev sahibinin kazandığı maçlar **% 89,7**
- Beraberlikle biten **% 0.3**,
- Deplasman takımının kazandığı maçlar ise **% 33,6** oranında tahminlenebilmiştir.

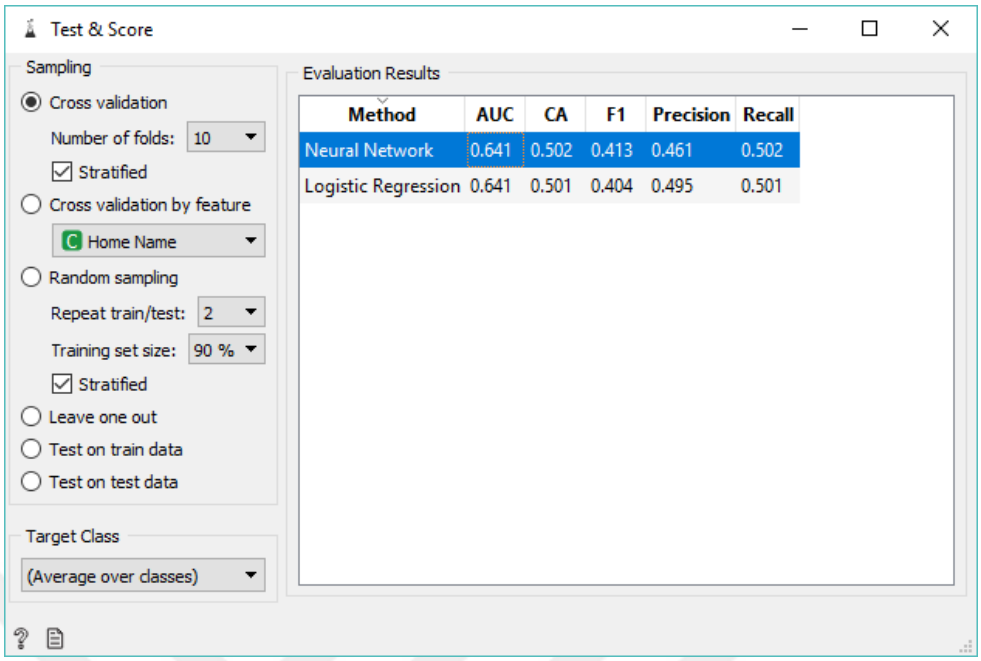
6.2 Neural Network (Sinir Ağları) & Başarı Oranları

Learner hiperparametreleri Şekil 6.3'deki gibi seçilmiştir. Tek katmanlı yapıda bu parametrelerin farklı kombinasyonlarının sonuçlarda önemli bir değişikliği gözlenmemiştir.



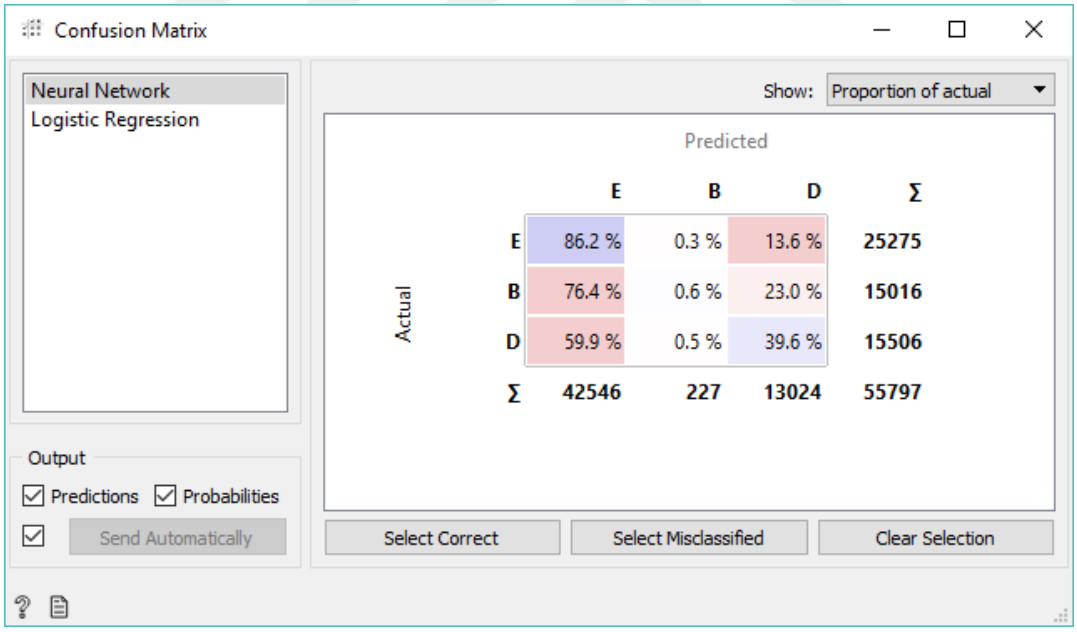
Şekil 6. 3 İddaa modeli ann hiperparametre seçimi

Öğrenme ve tahminleme işlemi sonrası Şekil 6.4'deki sonuçlar ile karşılaşılmıştır:



Şekil 6. 4 İddaa modeli ann sonuçları

Aranan sonuç için ortalama doğruluk %50 seviyesindedir.



Şekil 6. 5 İddaa modeli ann hata matrisi

Sınıf değerleri bazındaki doğruluk değerleri Şekil 6.5'deki gibi bulunmuştur.

- Ev sahibinin kazandığı maçlar **% 86,2**
- Beraberlikle biten **% 0.6**,
  Deplasman takımının kazandığı maçlar ise **% 39,6** oranında tahminlenebilmiştir.

7
GÜRÜLTÜ AZALTMA

Gürültü: Parametrenin kendisini doğru şekilde temsil etmeyen, parametre ve sonuç arasında herhangi bir ilişki sağlamayani istenmeyen veridir.

Gürültü Tipleri

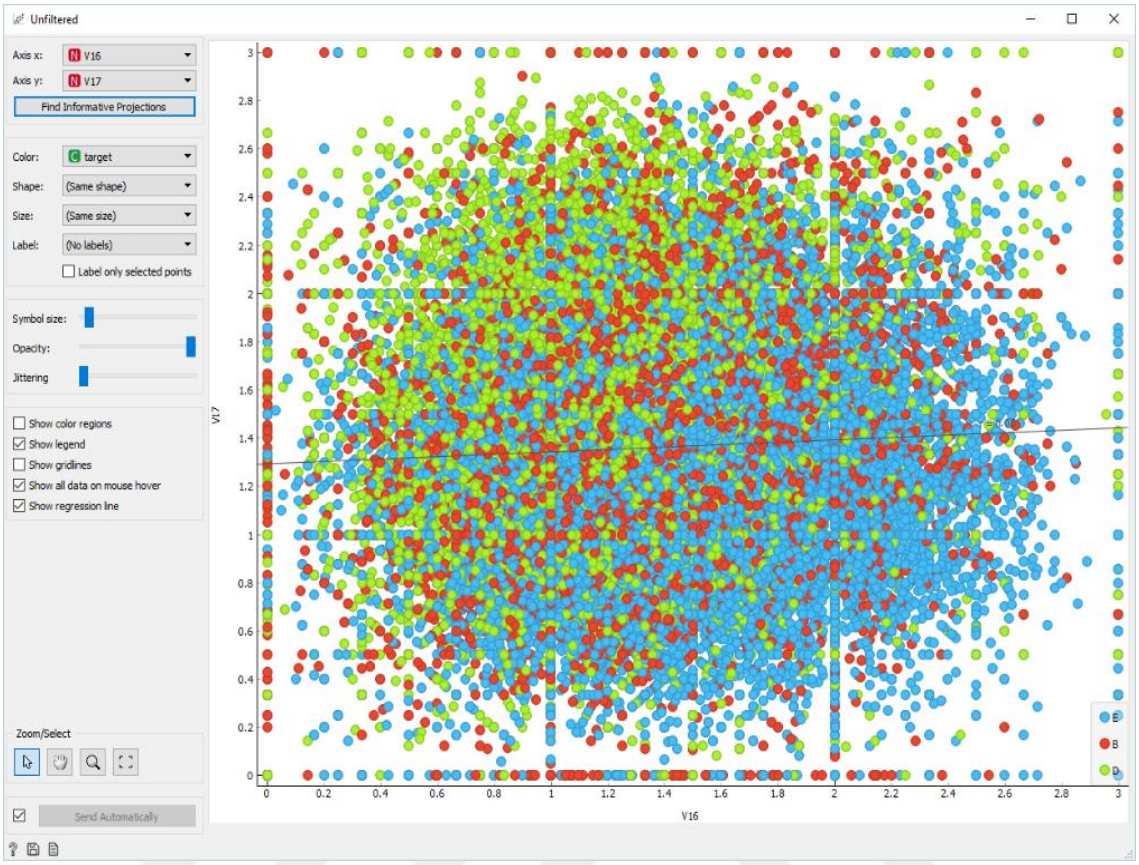
- Parametre ve Hedef değerlerindeki anomaliteler
- Hedef değeri anlamaya yardımcı olmayan parametreler
- Belirli bir formu veya ilişkisi olmayan kayıtlar

Gürültünün azaltılmasının etkileri

- Algoritmanın daha hızlı elde edilmesini sağlar
- Modeli basitleştirir ve modelin sunumunu rahatlatır
- Doğru parametre kümesi seçilirse (gürültü olan parametrenin kaldırılması) modelin doğruluğunu arttırır
- Overfitting'i engeller

V16, V17 parametrelerinin incelenmesi

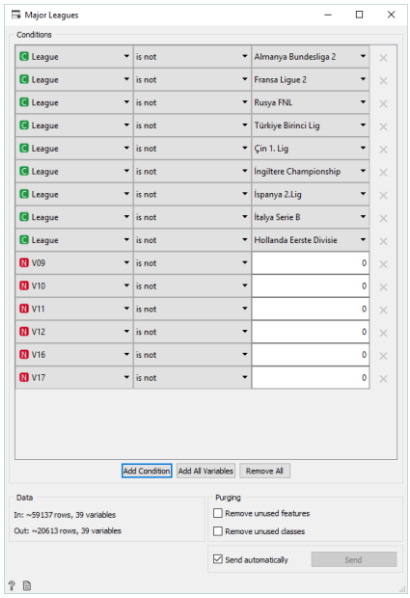
Bu iki parametre için saçılım grafiğine bakıldığında beklenen ilişkiyi bozan değerler olduğu farkedilmiştir. Bu değerler incelendiğinde bazı Küçük liglerdeki verilerde eksiklikler olduğu farkedilmiştir. 0 değerinde yığılma olduğu görülmüştür.



Şekil 7. 1 V16 – v17 parametreleri saçılım grafiği

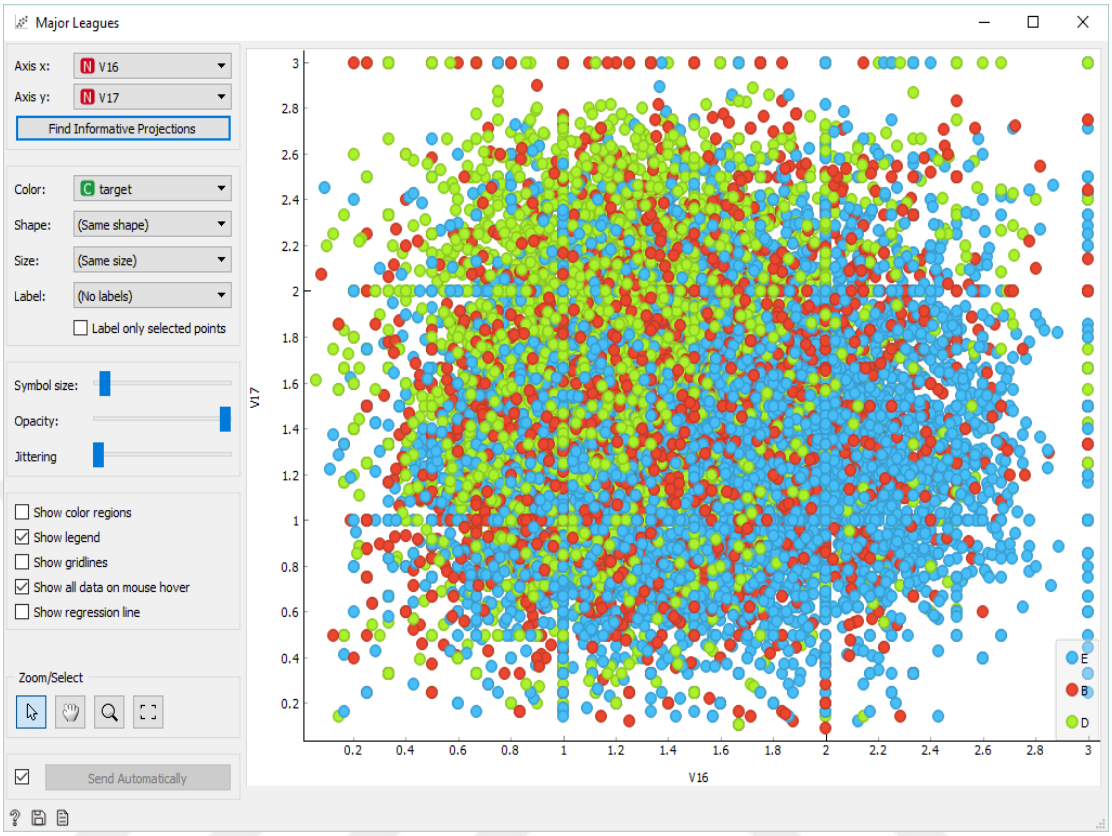
Öncelikle verisetinin iki ayrı parçaya ayrılması ve incelemenin bu şekilde devam edilmesine karar verilmiştir.

Büyük Ligler: Veri setinden 2. ligler çıkarılmıştır.



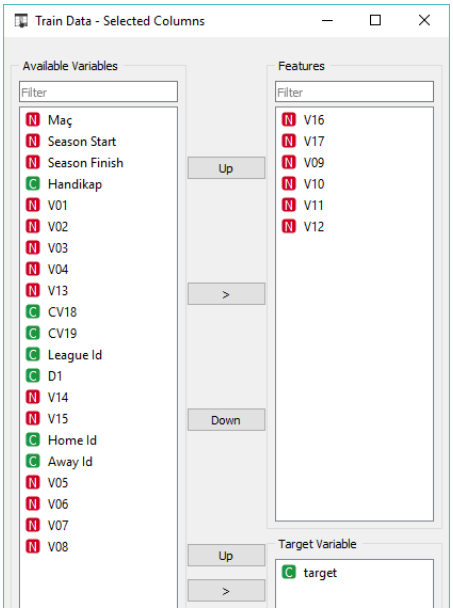
Şekil 7. 2 Büyük lig seçimi

Bu filtreleme işlemi sonrası saçılım grafiğinde iyileşme gözlenmiştir. 0'daki yığılma düzelmiştir.



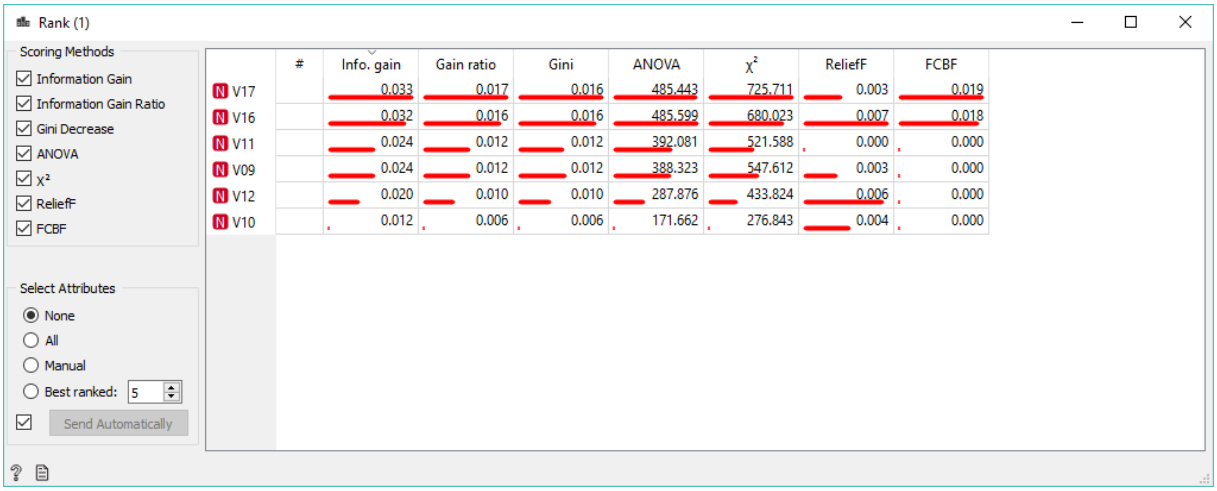
Şekil 7. 3 V16 – v17 parametreleri saçılım grafiği - 2

Modelin geliştirilmesi için Şekil 7.4'deki parametreler seçilmiştir:



Şekil 7. 4 Büyük lig modeli parametre seçimi

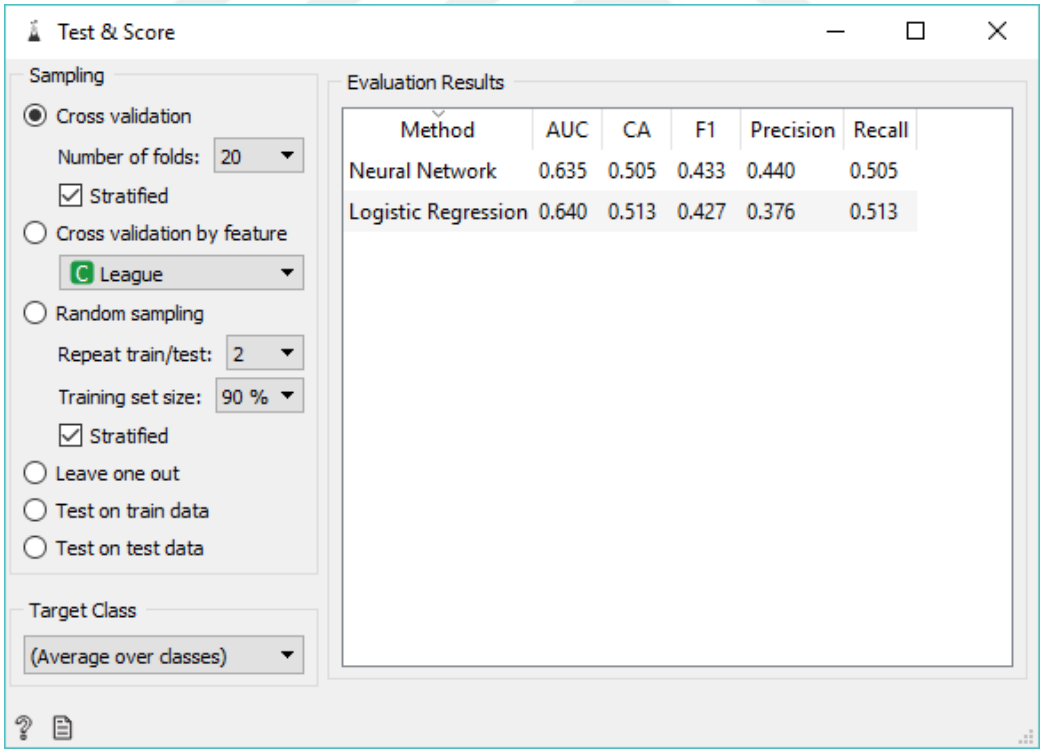
Bu parametreler için skorlama skorları Şekil 7.5'deki gibi gözlenmiştir:



Şekil 7. 5 Büyük lig modeli skorlama

En yüksek skorlar V17 ve V16 parametreleri için elde edilmiştir.

Cross validation (Çapraz Doğrulama) ile sonuçlar Şekil 7.6'daki şekilde elde edilmiştir.

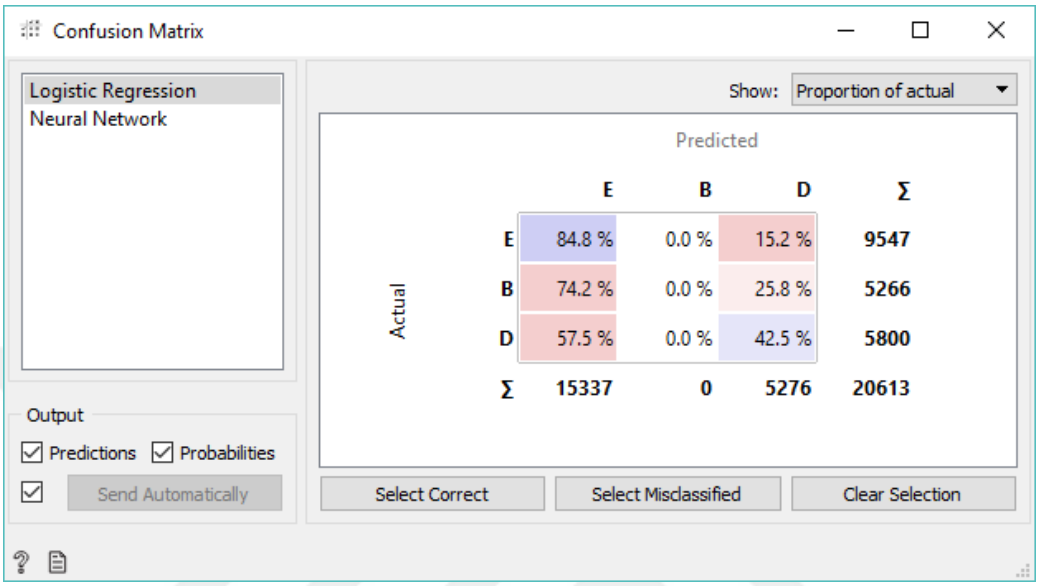


Şekil 7. 6 Büyük lig modeli cross validation (çapraz doğrulama) sonuçları

Neural Network (Sinir Ağları) modelinin tahminlerinin başarı oranı %44, gerçekleşen sonucu bilme oranı %50,5 olarak hesaplanmıştır.

Lojistik regresyon modelinin tahminlerinin başarı oranı %37,6, gerçekleşen sonucu bilme oranı %51,3 olarak hesaplanmıştır.

7.1 Lojistik Regresyon & Başarı Oranları



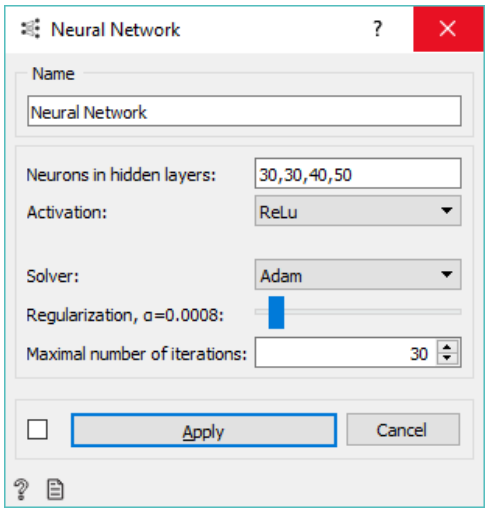
Şekil 7. 7 Büyük lig modeli lr hata matrisi

Sınıf değerleri bazındaki doğruluk değerleri Şekil 7.7'deki gibi bulunmuştur.

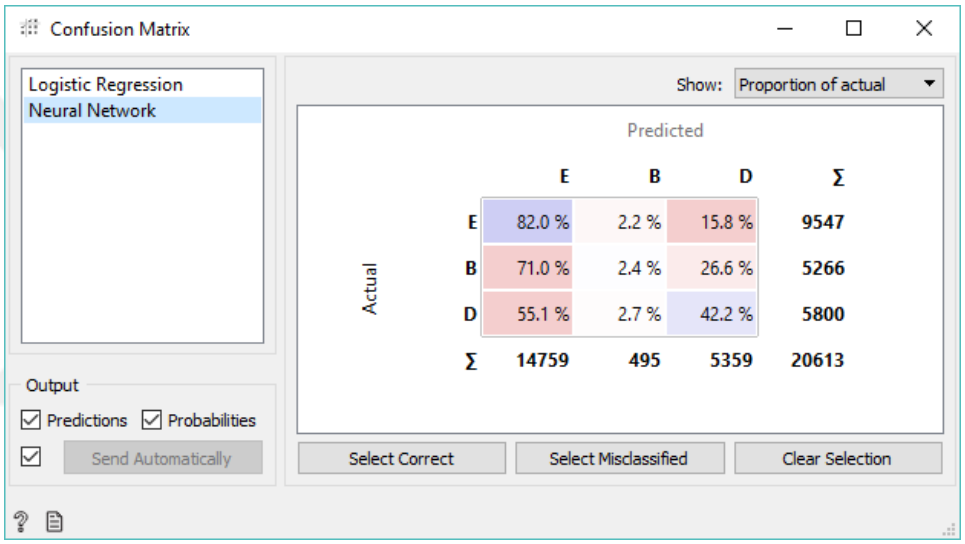
- Ev sahibinin kazandığı maçlar **% 84,8**
- Beraberlikle biten **% 0**,
- Deplasman takımının kazandığı maçlar ise **% 42,5** oranında tahminlenebilmiştir.

7.2 Neural Network (Sinir Ağları) & Başarı Oranları

Aktivasyon fonksiyonu olarak Şekil 7.8'deki gibi Relu seçilmiştir.



Şekil 7. 8 Büyük lig modeli ann hiperparametre seçimi



Şekil 7. 9 Büyük lig modeli ann hata matrisi

Sınıf değerleri bazındaki doğruluk değerleri Şekil 7.9'daki gibi bulunmuştur.

- Ev sahibinin kazandığı maçlar **% 82,**
- Beraberlikle biten **% 2.4**,
- Deplasman takımının kazandığı maçlar ise **% 42,2** oranında tahminlenebilmiştir

8
İKİNCİ MODEL

Deney 1 (Uzun Dönem Katsayı)

Tanımlar

Uzun Dönem Katsayı: Uzun dönem takımların itibarı/performansı

Hesaplama Yöntemi

Parametre setindeki HomeId ve AwayId kategorize değerlerinin modele katkı sağlayabilmesi amacıyla takımları temsil edecek Uzun Dönem Performansı için bir metrik planlanmıştır.

Aktif sezondan önceki tüm sezonlarda kazanılan puanların maç sayısına oranı şeklinde hesaplanacaktır.

Eğer önceki sezon/larda hesaplanacak bir değer yoksa bu alana -1 değeri basılacaktır. Böylece ilgili takım, mevcut lige yeni katılmış olarak işaretlenecektir

Parametreler

LTR_H Ev sahibi takımın uzun dönem performansı

LTR_A Deplasman takımın uzun dönem performansı

parametreleri eklenmiştir.

LTR_H Tanımsız ise -1

LTR_A Tanımsız ise -1

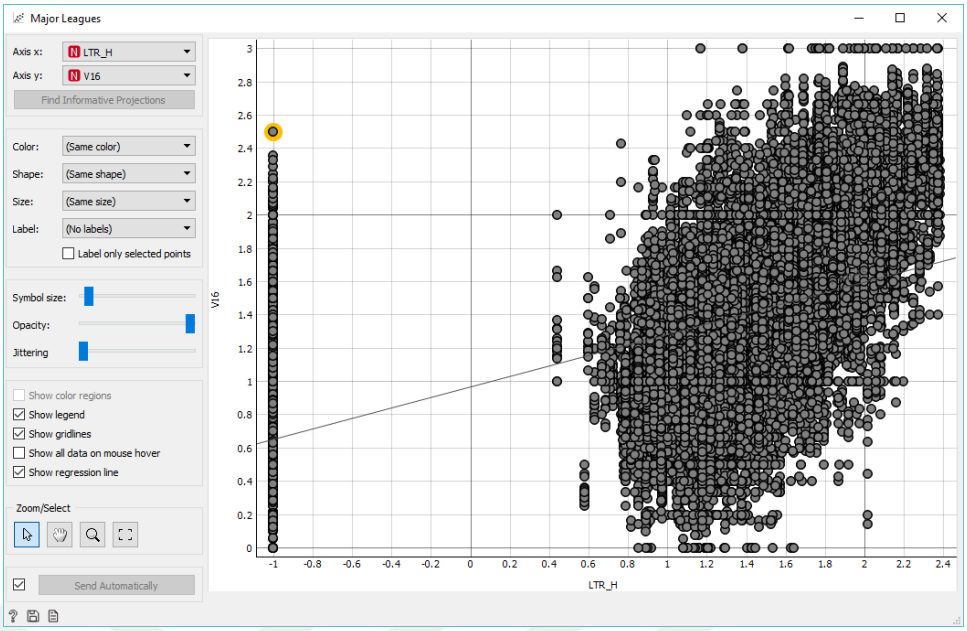
kuralları belirlenmiştir.

Büyük Ligler seçilmiştir.

Parametreler için herhangi bir filtre tanımlanmamıştır.

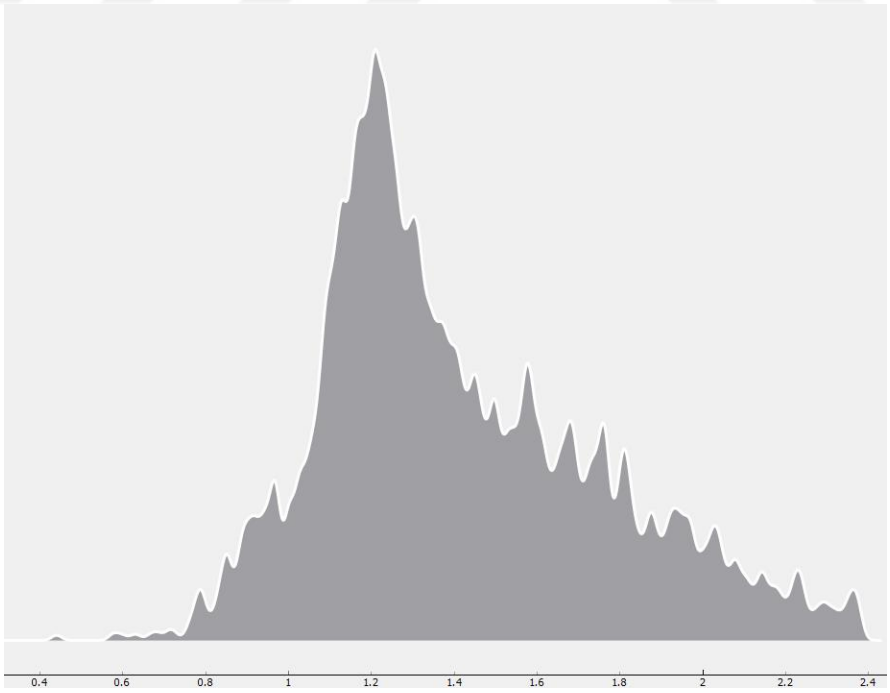
LTR Parametresinin incelenmesi

Tanımsız değerler filtrelenmeden incelendiğinde Şekil 8.1'deki saçılım grafiği ile karşılaşılmıştır.



Şekil 8. 1 Büyük lig modeli ltr saçılımı

Bu grafikten bir başka bir ligden gelen takımların (-1 değerli) ligde üst sıralar tırmanabildikleri gözlenmiştir. LTR değeri -1 olup V16 parametresi 2,5 olan takım var. Genel LTR dağılımı Şekil 8.2'deki gibidir;



Şekil 8. 2 Büyük lig modeli ltr dağılımı

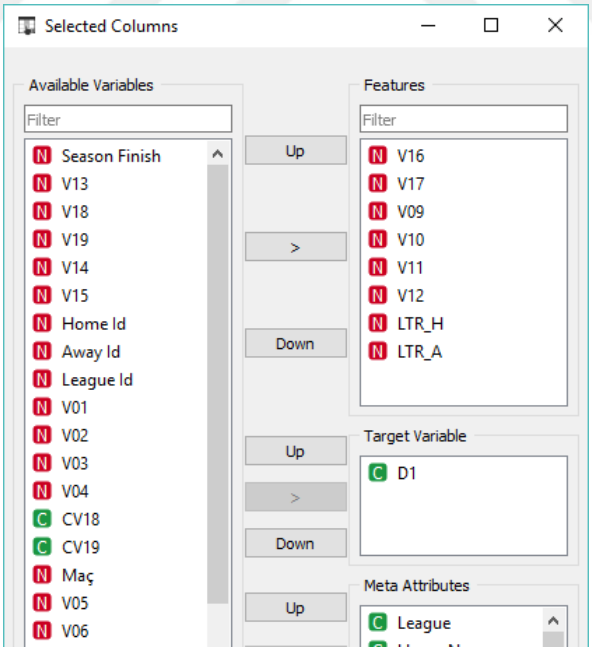
1,2 üzerinde yığılma gözlenmiştir.

Ayrıca bu parametre seti için skorlama yapılmış ve Şekil 8.3'deki değerler ile karşılaşılmıştır:

Şekil 8. 3 Büyük lig ltr modeli skorlama

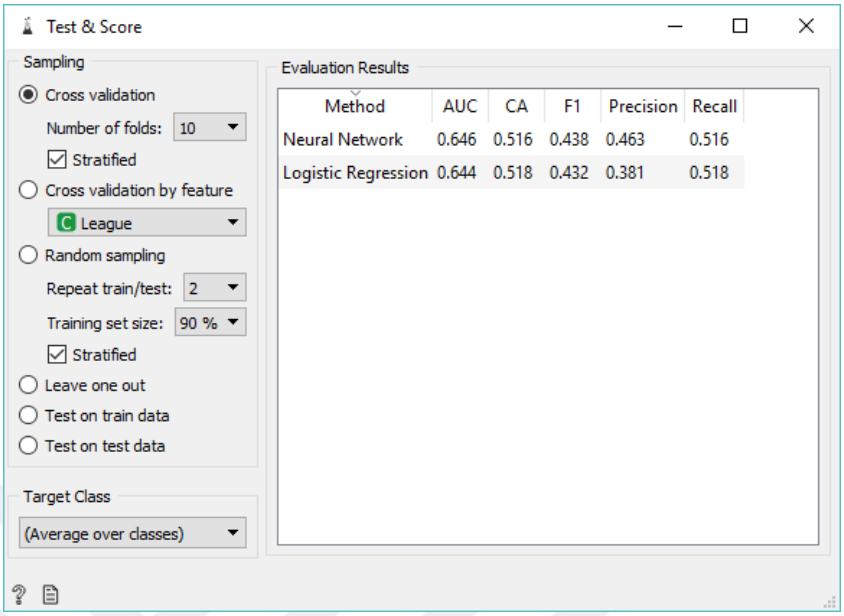
Bu parametre setinin sonuç ile kuvvetli bir ilişki içerisinde olduğu gözlenmektedir.

Modelin geliştirilmesi için Şekil 8.4'deki parametreler seçilmiştir:



Şekil 8. 4 Büyük lig ltr modeli parametre seçimi

Cross validation (Çapraz Doğrulama) ile sonuçlar Şekil 8.5'deki şekilde elde edilmiştir.



Şekil 8. 5 Büyük lig ltr modeli sonuçları

Neural Network (Sinir Ağları) modelinin tahminlerinin başarı oranı %46,3, gerçekleşen sonucu bilme oranı %51,6 olarak hesaplanmıştır.

Lojistik regresyon modelinin tahminlerinin başarı oranı %38,1, gerçekleşen sonucu bilme oranı %51,8 olarak hesaplanmıştır.

Sonuç

Bu parametre seti HomeId ve AwayId 'i doğrudan kullanmak yerine takımların kategorize edilebilmesini sağlamıştır.

LTR_H ve LTR_A değerleri için olduğu kadar V09, V10, V11 ve V12 parametreleri için de tanımsız değerin -1 yapılması, verisetinin incelenmesini kolaylaştıracak ve ayrıştırma işlemine yardımcı olacaktır.

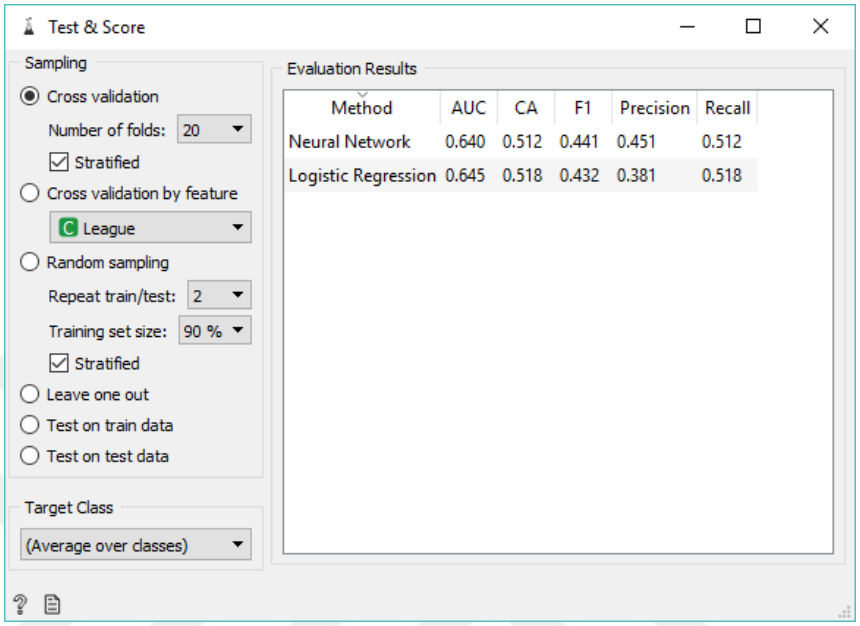
Deney 2 (V9-V12 - Tanımsız Değer Ataması)

Bu deney, V9 – V12 parametrelerinin tanımsız değerinin değiştirilmesi işlemi ile ilgili hazırlanmıştır.

Hesaplama Yöntemi

Eğer önceki sezon/larda hesaplanacak bir değer yoksa bu alana -1 değeri basılacaktır. Böylece ilgili takım, mevcut lige yeni katılmış olarak işaretlenecektir.

Cross validation (Çapraz Doğrulama) ile sonuçlar Şekil 8.6'daki şekilde elde edilmiştir.



Şekil 8. 6 Büyük lig tanımsız değer modeli sonuçları

Neural Network (Sinir Ağları) modelinin tahminlerinin başarı oranı %45,1, gerçekleşen sonucu bilme oranı %51,2 olarak hesaplanmıştır.

Lojistik regresyon modelinin tahminlerinin başarı oranı %38,1, gerçekleşen sonucu bilme oranı %51,8 olarak hesaplanmıştır.

Sonuç

V09, V10, V11 ve V12 parametreleri için tanımsız değerin -1 yapılması, verisetinin incelenmesini kolaylaştırmış ve ayrıştırma işlemine yardımcı olmuştur.

Deney 3 (Kısa Dönem Son X - Karşılıklı Gol)

Bu rapor, STLX (Kısa Dönem Son X : Kısa dönem kabul edilen son X maç) adında yeni bir parametrenin verisetine eklenip eklenemeyeceği ile ilgili yapılan çalışmaları kapsamaktadır.

Hesaplama Yöntemi

Ev sahibi/Deplasman takımın son X maçı için aşağıdaki formül uygulanacaktır.

Örnek:

HomeTeam = Valencia,

Away_Team = Osasuna

HomeTeam 'in önceki haftalardaki son X maçında

HomeRate = Ev sahibi takımın attığı gol sayısı x rakip takımın V17 değeri

AwayRate = Deplasman takımın attığı gol sayısı x ev sahibi takımın V16 değeri

(Not: Böylece Barcelonaya gol atmak ile Villareale gol atmak farklı hale gelecek)

Eğer HomeTeam (Valencia) bu maçta ev sahibi ise, (örneğin, Valencia - Atletico Madrid)

HomeRate - AwayRate

deplasmanda ise, (örneğin, Mallorca - Valencia)

AwayRate - HomeRate hesaplanacak.

Son 3 maç için bu değerlerin ortalaması alınacak ve STL3_H parametresi elde edilecek.

Aynı işlem deplasman takımı (Osasuna) için de yapıldı ve STL3_A adı verilecek.

Parametreler

STL3_H	Son 3 maçtaki gol performansı (V16, V17 ağırlıklı)
STL3_A	Son 3 maçtaki gol performansı (V16, V17 ağırlıklı)
STL2_H	Son 2 maçtaki gol performansı (V16, V17 ağırlıklı)
STL2_A	Son 2 maçtaki gol performansı (V16, V17 ağırlıklı)
STL1_H	Son 1 maçtaki gol performansı (V16, V17 ağırlıklı)
STL1_A	Son 1 maçtaki gol performansı (V16, V17 ağırlıklı)

Parametreleri de eklendi.

STL3_H	Tanımsız ise -9999
STL3_A	Tanımsız ise -9999
STL2_H	Tanımsız ise -9999
STL2_A	Tanımsız ise -9999
STL1_H	Tanımsız ise -9999
STL1_A	Tanımsız ise -9999

Tanımlamaları yapılmıştır.

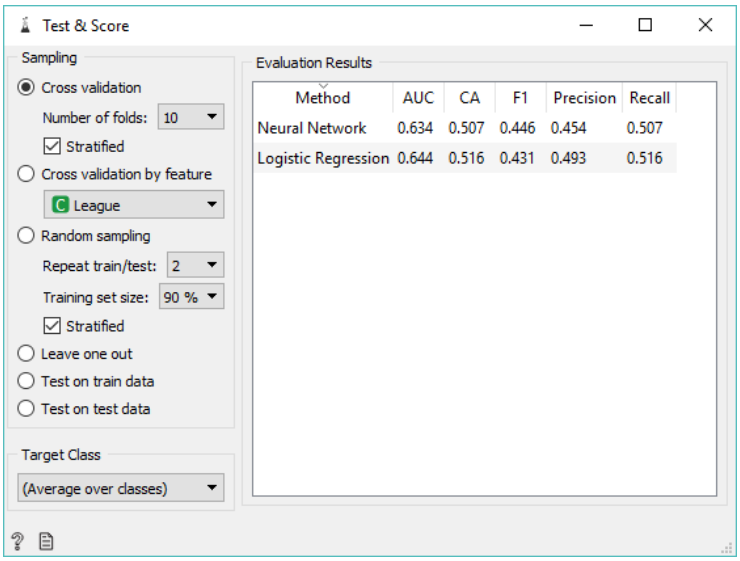
Skor Şekil 8.7'deki gibi gözlenmiştir;

	#	Info. gain	Gain ratio	Gini	ANOVA	χ^2	ReliefF	FCBF
LTR_A		0.039	0.020	0.019	423.311	1141.364	0.001	0.023
LTR_H		0.037	0.018	0.018	432.707	1086.965	0.002	0.021
V16		0.034	0.017	0.017	743.946	939.691	0.005	0.000
V17		0.032	0.016	0.016	704.106	1007.195	0.010	0.000
V09		0.028	0.014	0.014	445.389	849.106	0.003	0.000
V11		0.025	0.013	0.012	404.291	736.777	0.000	0.000
V05		0.021	0.010	0.010	445.870	609.201	-0.004	0.000
V07		0.017	0.008	0.008	343.716	509.203	0.006	0.000
V06		0.011	0.006	0.006	197.028	371.277	0.001	0.000
V08		0.011	0.006	0.006	211.982	384.616	0.001	0.000
V12		0.007	0.004	0.004	9.783	148.700	0.001	0.000
V10		0.005	0.003	0.003	32.164	54.284	0.001	0.000
STL3_H		0.003	0.002	0.002	40.674	60.372	0.004	0.000
STL3_A		0.003	0.001	0.001	36.775	39.069	0.001	0.000
STL2_H		0.002	0.001	0.001	1.385	41.753	0.000	0.000
STL1_H		0.002	0.001	0.001	0.917	8.826	0.000	0.000
STL2_A		0.001	0.001	0.001	3.154	22.515	0.000	0.000
STL1_A		0.001	0.001	0.001	1.227	5.891	-0.000	0.000

Şekil 8. 7 Büyük lig kısa dönem modeli skorlama

En yüksek skorlar LTR parametreleri için elde edilmiştir.

Cross validation (Çapraz Doğrulama) ile sonuçlar Şekil 8.8'deki şekilde elde edilmiştir;



Şekil 8. 8 Büyük lig kısa dönem modeli sonuçları

Neural Network (Sinir Ağları) modelinin tahminlerinin başarı oranı %45,4, gerçekleşen sonucu bilme oranı %50,7 olarak hesaplanmıştır.

Lojistik regresyon modelinin tahminlerinin başarı oranı %49,3, gerçekleşen sonucu bilme oranı %51,6 olarak hesaplanmıştır.

9
MODEL PARAMETRİZASYONU

9.1 Liglerin Sınıflandırılması

Daha önceki çalışmalarda, verisetindeki toplam 19 lige ait maçların iki ayrı verisetine bölünmesine karar verilmişti.

Bu kapsamada iki ana lig sınıfı olacaktır:

Büyük Lig

Ülkelerin en üst seviyeli liglerini içerecektir:

1. Almanya Bundesliga
2. Belçika Lig A
3. Fransa Ligue 1
4. Hollanda Eredivisie
5. Rusya Premier Lig
6. Türkiye Süper Lig
7. Çin Süper Lig
8. İngiltere Premier League
9. İspanya Primera Lig
10. İtalya Serie A

Küçük Lig

Ülkelerin bir alt liglerini içerecektir:

1. Almanya Bundesliga 2
2. Fransa Ligue 2
3. Hollanda Eerste Divisie
4. Rusya FNL
5. Türkiye Birinci Lig
6. Çin 1. Lig
7. İngiltere Championship
8. İspanya 2.Lig
9. İtalya Serie B

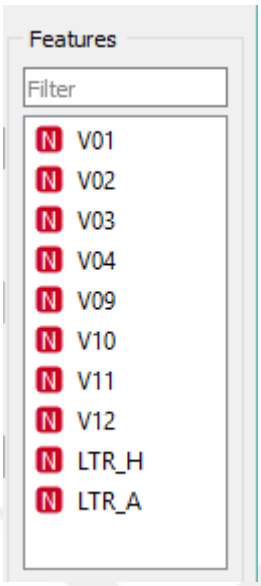
9.2 Hafta Numarasına Göre Sınıflandırma

Sezon arası antreman düzeyi, motivasyon, takımlardaki oyuncu değişikliği, üst lige çıkma, alt lige düşme vb benzeri değişiklikler sebebi ile futbol sezonundaki ilk maçların o takımların gerçek performansını yansıtmadığı bilinmektedir.

Bu bakımdan, ligin ilk haftalarına ait model kurgulanırken aktif sezona ait parametrelerin kullanılması, takımların gerçek durumlarını temsil etmeyecek bir başka söylemle, tahmin modellerimizde hataya sebebiyet verecektir. Bu bakımdan iki ayrı sınıflandırma modeli öngörülmüştür;

Geçmiş model: Bu model ligin ilk 4 haftasındaki tahminler için kullanılacak ve geçmişe ait takımların genel durumlarını ortaya koyacaktır. Şekil 9.1'deki parametreler öngörülmüştür.

Aktif sezon modeli: Bu model 5. hafta ve sonrası için kullanılacak modeldir.

Geçmiş	Aktif sezon
	

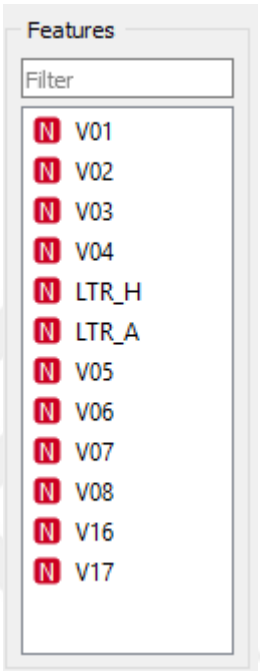
Şekil 9. 1 Geçmiş ve aktif sezon parametreleri

Fark edildiği üzere, V05 – V08, V16 – V17 içinde bulunan sezonla ilgili parametrelerdir.

9.3 Neural Network (Sinir Ağları) Modeli

Beraberlik oranlarının daha başarılı bilinebildiği modeldir.

Büyük ve Küçük ligler'de farklı parametreler kullanılacaktır

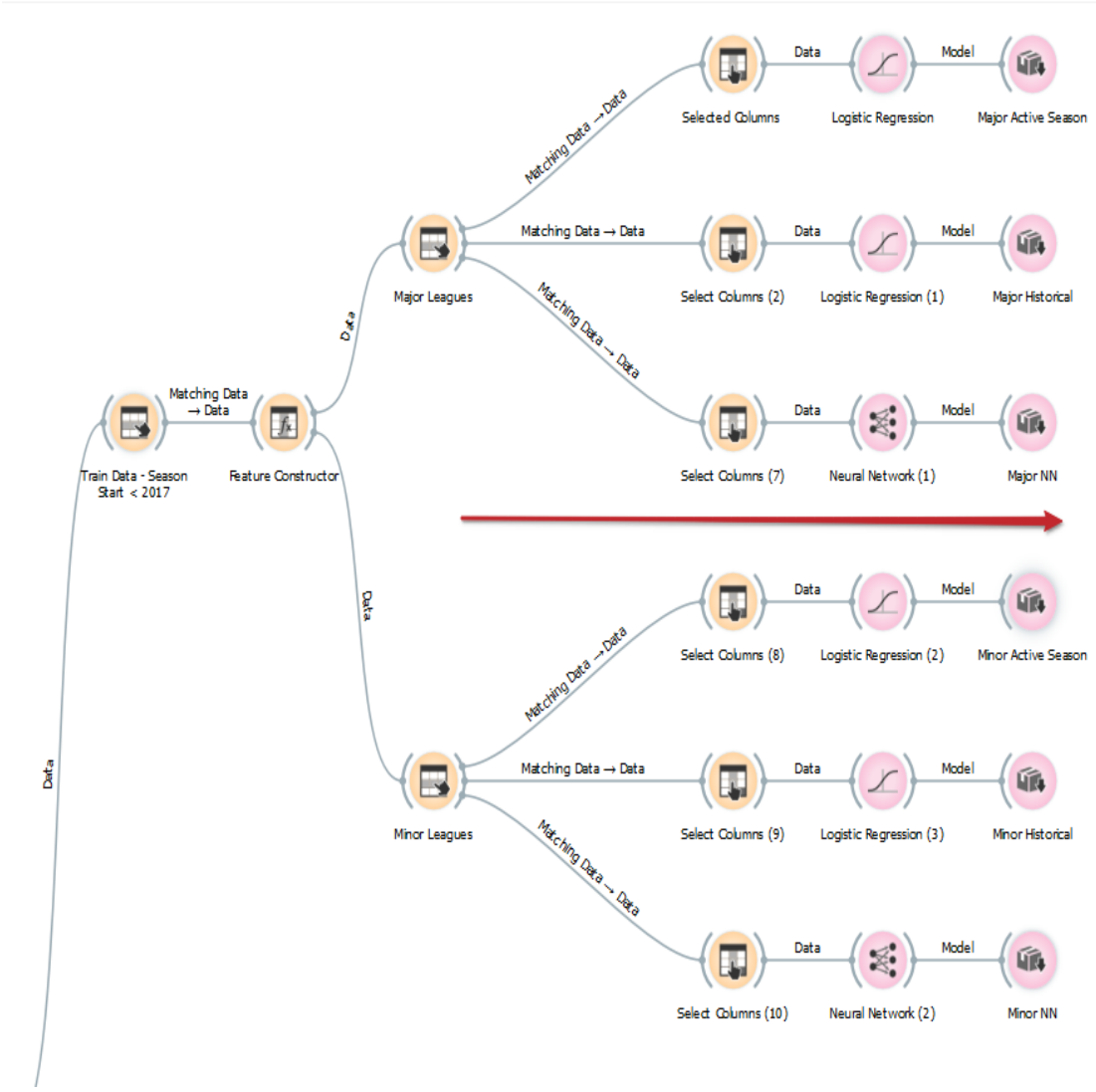
Büyük ligler	Küçük ligler
	

Şekil 9. 2 Büyük ve küçük lig parametreleri

Küçük liglerin yapısı gereği bir önceki yılın mevcut durum üzerinde çok da belirleyici olmaması sebebi ile V09-V12 parametreleri deneyler sonucunda çıkartılmıştır.

9.4 Modelin Öğrenilmesi Ve Saklanması İlk Aşama

Şekil 9.3'de görüldüğü üzere, ilk olarak parametre setinden 2017 öncesi veriler eğitim verisi olarak alınmıştır. Bu veriler önce büyük lig ve küçük lig olma durumlarına, sonra tahmin yapılan dönemin ilk 5 hafta veya sonrası olması durumuna göre ayrıştırılmış, lojistik regresyon ve sinir ağları için bu farklı durumlar dikkate alınarak modeller çalıştırılmıştır.



Şekil 9. 3 Modelin öğrenilmesi aşaması

9.5 Modelin Saklanma Altyapısı

Python temelli Flask adlı bir microframework ile python modellerine API desteği sağlanacaktır.

Tahminlemesi yapılacak instance(lar) API'ye yollanacak, flask bu API üzerinden aldığı parametreler için yukarıda anlatılan sınıflandırmalara göre ilgili modeli seçecek ve tahmin sonucunu cevap olarak geri dönecektir.

10
CANLI SKOR MODELİ (MAÇ İÇİ VERİLER)

Bu model, canlı maç skoru üzerinden yapay zeka modeli oluşturulması ile ilgilidir. Maç içi olayları dakika ve olay bazında elimizde mevcut ise bu bilgiler kullanılarak maç sırasında kullanılabilecek bir yapay zeka modeli oluşturulması hedeflenmiştir.

10.1 Parametreler

Çalışmanın ilerleyen safhalarında kullanılmak üzere maç sonu istatistikleri kullanılarak 1 grup oluşturulmuştur, Tablo 10.1'de gösterilmiştir.

Tablo 10. 1 Canlı skor modeli parametreler

ID	Analiz Tipi	Parametre
1	H2H (Head To Head)	Ev sahibi takımın, evinde yapılan maçlarda attığı gol ort
2	H2H (Head To Head)	Deplasman takımın, ev sahibi takımın evinde yapılan maçlarda attığı gol ort
3	H2H (Head To Head)	Deplasman takımın, evinde yapılan maçlarda attığı gol ort
4	H2H (Head To Head)	Ev sahibi takımın, deplasman takımın evinde yapılan maçlarda attığı gol ort
5	Gol Analizi	Ev Sahibi Takımın Bu Sezon Evinde Attığı Gol Sayısı Ortalaması
6	Gol Analizi	Ev Sahibi Takımın Bu Sezon Evinde Yediği Gol Sayısı Ortalaması
7	Gol Analizi	Deplasman Takımının Bu Sezon Deplasmanda Attığı Gol Sayısı Ortalaması
8	Gol Analizi	Deplasman Takımının Bu Sezon Deplasmanda Yediği Gol Sayısı Ortalaması
9	Gol Analizi	Ev Sahibi Takımın 1 Önceki Sezon Evinde Attığı Gol Sayısı Ortalaması
10	Gol Analizi	Ev Sahibi Takımın 1 Önceki Sezon Evinde Yediği Gol Sayısı Ortalaması
11	Gol Analizi	Deplasman Takımının 1 Önceki Sezon Deplasmanda Attığı Gol Sayısı Ortalaması
12	Gol Analizi	Deplasman Takımının 1 Önceki Sezon Deplasmanda Yediği Gol Sayısı Ortalaması
13	Gol Analizi	Ligde Bu Sezon Atılan Gol Sayısının Ortalaması
14	Lig Performansı – Son X	Ev Sahibi Takımın Son 5 Maçtaki Puan Oranı
15	Lig Performansı – Son X	DeplasmanTakımın Son 5 Maçtaki Puan Oranı
16	Lig Performansı	Ev Sahibi Takımın Maç Başına Puan Ortalaması
17	Lig Performansı	Deplasman Takımının Maç Başına Puan Ortalaması
18	Lig Performansı	Ev Sahibi Takımın Ligdeki Sıralamasının Normalize Değeri
19	Lig Performansı	Deplasman Takımının Ligdeki Sıralaması Normalize Değeri
20	Bahis Oranları	İddaa Maç Sonucu 1 Oranı
21	Bahis Oranları	İddaa Maç Sonucu 0 Oranı
22	Bahis Oranları	İddaa Maç Sonucu 2 Oranı
23	Skor	Ev Sahibi Takım Maç Sonucu Skoru
24	Skor	Deplasman Takımı Maç Sonucu Skoru
LTR_H		Ev sahibi takımın uzun dönem performansı
LTR_A		Deplasman takımın uzun dönem performansı

10.2 Yöntem

10.2.1 Maç Sonu İstatiskleri İle Alakalı Veri Setinin Model Kurgusundan Önce İşlenmesi

Yapılan incelemeler sonrasında Vol1 parametre seti için Tablo 10.2'deki sınır değerleri ön görülmüştür.

Tablo 10. 2 Canlı skor modeli parametreleri min max değerleri

ID	Min	Max	Not
1	0	7	5 maç yapmamışlarsa, kaç maç varsa o kadar maçta gol ort
2	0	7	5 maç yapmamışlarsa, kaç maç varsa o kadar maçta gol ort
3	0	7	5 maç yapmamışlarsa, kaç maç varsa o kadar maçta gol ort
4	0	7	5 maç yapmamışlarsa, kaç maç varsa o kadar maçta gol ort
5	0	7	Toplam İçin Max 100
6	0	7	Toplam İçin Max 100
7	0	7	Toplam İçin Max 100
8	0	7	Toplam İçin Max 100
9	0	7	Toplam İçin Max 100 – Tanımsız ise -1
10	0	7	Toplam İçin Max 100 – Tanımsız ise -1
11	0	7	Toplam İçin Max 100 – Tanımsız ise -1
12	0	7	Toplam İçin Max 100 – Tanımsız ise -1
13	0	7	
14	0	1	Daha 5 maç oynanmamışsa alınan puan/oynanan maç sayısı*3
15	0	1	Daha 5 maç oynanmamışsa alınan puan/oynanan maç sayısı*3
16	0	3	Tanımsız ise -1
17	0	3	Tanımsız ise -1
18	0	1	Tanımsız ise -1
19	0	1	Tanımsız ise -1
LTR_H			Tanımsız ise -1
LTR_A			Tanımsız ise -1
STL3_H			Tanımsız ise -20
STL3_A			Tanımsız ise -20
STL2_H			Tanımsız ise -20
STL2_A			Tanımsız ise -20
STL1_H			Tanımsız ise -20
STL1_A			Tanımsız ise -20

10.2.2 Maç İçi Olaylardan Bir Veri Setinin Elde Edilmesi

Veritabanındaki maç içi (match event) olay listesi aşağıdaki gibidir:

'GOAL', 'GOAL_FROM_PENALTY', 'OWN_GOAL', 'YELLOW_CARD', 'RED_CARD', 'YELLOW_RED_CARD', 'SUBSTITUTED', 'MISSED_PENALTY'.

Bu veriler işlenerek maç bazında 0. Ve 92. Dakika (uzatma sonrası maç sonunu temsil edecektir) arasında dakikda ve belirli metrikler oluşturulmuştur. 90+ sonrasında kaç dakika oynanmışsa o dakikalardaki olaylar 91. Dakika altında toplanmıştır:

Buradaki amaç dakika ilgili metriklerin son durumu ve maç sonucu arasındaki ilişkinin araştırılmasıdır.

Parametre listesi ve açıklaması:

Tablo 10. 3 Parametre listesi

Parametre adı	**Açıklaması**
match_id	Maç id
Minute	Dakika
home_team_score	ev sahibi takım skoru
Score	deplasman takımın skoru
home_team_yellow_card	Ev sahibi takım toplam sarı kart
away_team_yellow_card	Deplasman takım toplam sarı kart
home_team_red_card	Ev sahibi takım toplam kırmızı kart
away_team_red_card	Deplasman takım toplam kırmızı kart
home_substitution	Ev sahibi toplam değişiklik sayısı
away_substitution	Deplasman takım değişiklik sayısı
home_missed_penalty	Ev sahibi takım kaçan penaltı sayısı
away_missed_penalty	Deplasman takı kaçan penaltı sayısı
home_team_ft_score	Ev sahibi takım maç sonu skor
away_ft_team_score	Deplasman takım maç sonu skor
score_difference	home_team_score - away_team_score
yellow_card_difference	home_team_yellow_card - away_team_yellow_card
red_card_difference	home_team_red_card - away_team_red_card
substitution_difference	home_substitution - away_substitution
winning_side	Maç sonu kazanan taraf

Dosya içeriğinden bir kesit:

match_id,minute,home_team_score,away_team_score,home_team_yellow_card,away_team_yellow_card,home_team_red_card,away_team_red_card,home_substitution,away_substitution,home_missed_penalty,away_missed_penalty,home_team_ft_score,away_ft_team_score,score_difference,yellow_card_difference,red_card_difference,substitution_difference,winning_side

9,0,0,0,0,0,0,0,0,0,0,0,1,2,0,0,0,0,2

9,1,0,0,0,0,0,0,0,0,0,0,1,2,0,0,0,0,2

9,2,0,0,0,0,0,0,0,0,0,0,1,2,0,0,0,0,2

9,3,0,0,0,0,0,0,0,0,0,0,1,2,0,0,0,0,2

9,4,0,1,0,0,0,0,0,0,0,0,1,2,-1,0,0,0,2

9,5,0,1,0,0,0,0,0,0,0,0,1,2,-1,0,0,0,2

9,6,0,1,0,0,0,0,0,0,0,0,1,2,-1,0,0,0,2

9,7,0,1,0,0,0,0,0,0,0,0,1,2,-1,0,0,0,2

9,8,0,1,0,0,0,0,0,0,0,0,1,2,-1,0,0,0,2

9,9,0,1,0,0,0,0,0,0,0,0,1,2,-1,0,0,0,2

9,10,0,1,0,0,0,0,0,0,0,0,1,2,-1,0,0,0,2

9,11,0,1,0,0,0,0,0,0,0,0,1,2,-1,0,0,0,2

9,12,0,1,0,0,0,0,0,0,0,0,1,2,-1,0,0,0,2

9,13,0,1,0,0,0,0,0,0,0,0,1,2,-1,0,0,0,2

9,14,0,1,0,0,0,0,0,0,0,0,1,2,-1,0,0,0,2

9,15,0,1,0,0,0,0,0,1,0,0,1,2,-1,0,0,-1,2

9,16,0,1,0,0,0,0,0,1,0,0,1,2,-1,0,0,-1,2

9,17,0,1,0,0,0,0,0,1,0,0,1,2,-1,0,0,-1,2

9,18,0,1,0,0,0,0,0,1,0,0,1,2,-1,0,0,-1,2

9,19,0,1,0,0,0,0,0,1,0,0,1,2,-1,0,0,-1,2

9,20,0,1,0,0,0,0,0,1,0,0,1,2,-1,0,0,-1,2

9,21,0,1,0,0,0,0,0,1,0,0,1,2,-1,0,0,-1,2

9,22,0,1,0,0,0,0,0,1,0,0,1,2,-1,0,0,-1,2

9,23,0,1,0,0,0,0,0,1,0,0,1,2,-1,0,0,-1,2

9,24,0,1,0,0,0,0,0,1,0,0,1,2,-1,0,0,-1,2

9,25,0,1,0,0,0,0,0,1,0,0,1,2,-1,0,0,-1,2

9,26,0,1,0,0,0,0,0,1,0,0,1,2,-1,0,0,-1,2

9,27,0,1,0,0,0,0,0,1,0,0,1,2,-1,0,0,-1,2

9,28,0,1,0,0,0,0,0,1,0,0,1,2,-1,0,0,-1,2

9,29,0,1,0,0,0,0,0,1,0,0,1,2,-1,0,0,-1,2

9,30,0,1,0,0,0,0,0,1,0,0,1,2,-1,0,0,-1,2

9,31,0,1,0,0,0,0,0,1,0,0,1,2,-1,0,0,-1,2

9,32,0,1,0,0,0,0,0,1,0,0,1,2,-1,0,0,-1,2

9,33,0,1,0,0,0,0,0,1,0,0,1,2,-1,0,0,-1,2

9,34,0,1,0,0,0,0,0,1,0,0,1,2,-1,0,0,-1,2

9,35,0,1,0,0,0,0,0,1,0,0,1,2,-1,0,0,-1,2

9,36,0,1,0,0,0,0,0,1,0,0,1,2,-1,0,0,-1,2

9,37,0,1,1,0,0,0,0,1,0,0,1,2,-1,1,0,-1,2

9,38,0,1,1,0,0,0,0,1,0,0,1,2,-1,1,0,-1,2

9,39,0,1,1,0,1,0,0,1,0,0,1,2,-1,1,1,-1,2

9,40,0,1,1,0,1,0,0,1,0,0,1,2,-1,1,1,-1,2

9,41,0,1,1,1,1,0,1,1,0,0,1,2,-1,0,1,0,2

9,42,0,1,1,1,1,0,1,1,0,0,1,2,-1,0,1,0,2

9,43,0,1,1,1,1,0,1,1,0,0,1,2,-1,0,1,0,2

9,44,0,1,1,1,1,0,1,1,0,0,1,2,-1,0,1,0,2

9,45,0,1,1,1,1,0,1,1,0,0,1,2,-1,0,1,0,2

9,46,0,1,1,1,1,0,1,1,0,0,1,2,-1,0,1,0,2

9,47,0,1,1,1,1,0,1,2,0,0,1,2,-1,0,1,-1,2

9,48,0,1,1,1,1,0,1,2,0,0,1,2,-1,0,1,-1,2

9,49,0,1,1,1,1,0,1,2,0,0,1,2,-1,0,1,-1,2

9,50,0,1,1,1,1,0,1,2,0,0,1,2,-1,0,1,-1,2

9,51,0,1,1,1,1,0,1,2,0,0,1,2,-1,0,1,-1,2

9,52,0,1,1,1,1,0,1,2,0,0,1,2,-1,0,1,-1,2

9,53,0,1,1,1,1,0,1,2,0,0,1,2,-1,0,1,-1,2

9,54,0,1,1,1,1,0,1,2,0,0,1,2,-1,0,1,-1,2

9,55,0,1,1,1,1,0,1,2,0,0,1,2,-1,0,1,-1,2

9,56,0,1,2,1,1,0,1,2,0,0,1,2,-1,1,1,-1,2

9,57,0,1,2,2,1,0,1,2,0,0,1,2,-1,0,1,-1,2

9,58,0,1,2,2,1,0,1,2,0,0,1,2,-1,0,1,-1,2

9,59,0,1,2,2,1,0,1,2,0,0,1,2,-1,0,1,-1,2

9,60,0,1,2,2,1,0,1,2,0,0,1,2,-1,0,1,-1,2

9,61,0,2,2,2,1,0,1,2,0,0,1,2,-2,0,1,-1,2

9,62,0,2,2,2,1,0,1,2,0,0,1,2,-2,0,1,-1,2

9,63,0,2,2,2,1,0,1,2,0,0,1,2,-2,0,1,-1,2

9,64,0,2,2,2,1,0,1,2,0,0,1,2,-2,0,1,-1,2

9,65,0,2,2,2,1,0,1,2,0,0,1,2,-2,0,1,-1,2

9,66,0,2,2,2,1,0,1,2,0,0,1,2,-2,0,1,-1,2

9,67,0,2,2,2,1,0,1,2,0,0,1,2,-2,0,1,-1,2

9,68,0,2,2,2,1,0,1,2,0,0,1,2,-2,0,1,-1,2

9,69,0,2,2,2,1,0,1,2,0,0,1,2,-2,0,1,-1,2

9,70,0,2,2,2,1,0,2,2,0,0,1,2,-2,0,1,0,2

9,71,0,2,2,2,1,0,2,2,0,0,1,2,-2,0,1,0,2

9,72,0,2,2,2,1,0,2,2,0,0,1,2,-2,0,1,0,2

9,73,0,2,2,2,1,0,2,3,0,0,1,2,-2,0,1,-1,2

9,74,0,2,2,2,1,0,2,3,0,0,1,2,-2,0,1,-1,2

9,75,0,2,2,2,1,0,2,3,0,0,1,2,-2,0,1,-1,2

9,76,0,2,2,2,1,0,2,3,0,0,1,2,-2,0,1,-1,2

9,77,0,2,2,2,1,0,3,3,0,0,1,2,-2,0,1,0,2

9,78,0,2,2,2,1,0,3,3,0,0,1,2,-2,0,1,0,2

9,79,1,2,2,2,1,0,3,3,0,0,1,2,-1,0,1,0,2

9,80,1,2,2,2,1,0,3,3,0,0,1,2,-1,0,1,0,2

9,81,1,2,2,2,1,0,3,3,0,0,1,2,-1,0,1,0,2

9,82,1,2,2,2,1,0,3,3,0,0,1,2,-1,0,1,0,2

9,83,1,2,2,2,1,0,3,3,0,0,1,2,-1,0,1,0,2

9,84,1,2,2,2,1,0,3,3,0,0,1,2,-1,0,1,0,2

9,85,1,2,2,2,1,0,3,3,0,0,1,2,-1,0,1,0,2

9,86,1,2,2,2,1,0,3,3,0,0,1,2,-1,0,1,0,2

9,87,1,2,2,2,1,0,3,3,0,0,1,2,-1,0,1,0,2

9,88,1,2,2,2,1,0,3,3,0,0,1,2,-1,0,1,0,2

9,89,1,2,2,2,1,0,3,3,0,0,1,2,-1,0,1,0,2

9,90,1,2,2,2,1,0,3,3,0,0,1,2,-1,0,1,0,2

9,91,1,2,2,2,1,0,3,4,0,0,1,2,-1,0,1,-1,2

9,92,1,2,2,2,1,0,3,4,0,0,1,2,-1,0,1,-1,2

Yukarıdaki örnekte 3. Dakikada rakip takım 1 gol atmış ve skor 0-1 olmuştur.

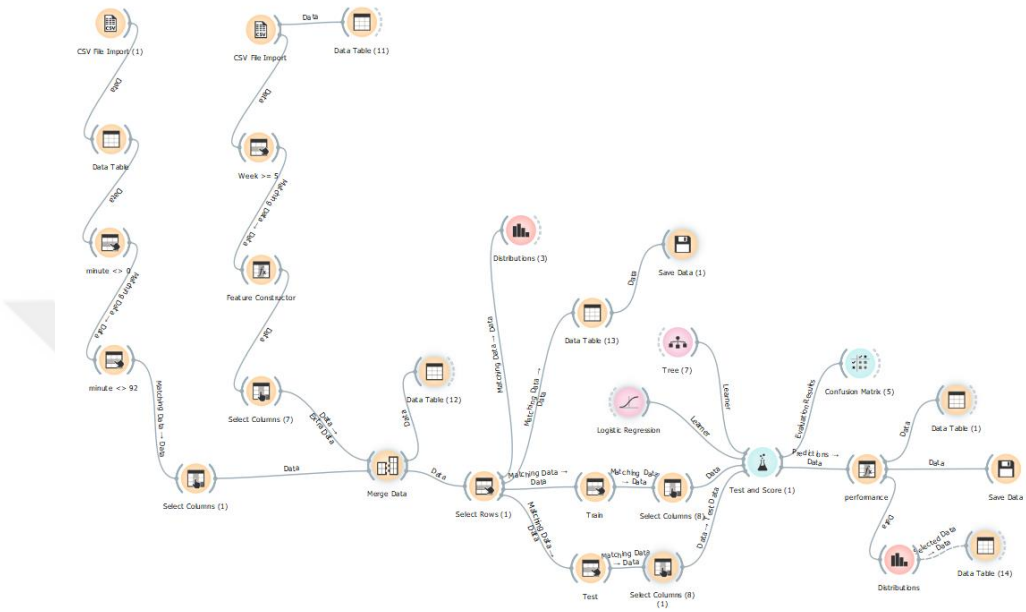
9,4,0,1,0,0,0,0,0,0,0,0,1,2,-1,0,0,0,2

Gene aynı şekilde 60.dakika içerisinde bir gol atışmıştır:

9,61,0,2,2,2,1,0,1,2,0,0,1,2,-2,0,1,-1,2

10.2.3 Sadece Maç Içi Verileri Kullanılarak Kazanan Tarafın Tahmin Edilmesi

Orange adlı yazılımda algortima Şekil 10.1'deki gibi kurgulanmıştır;



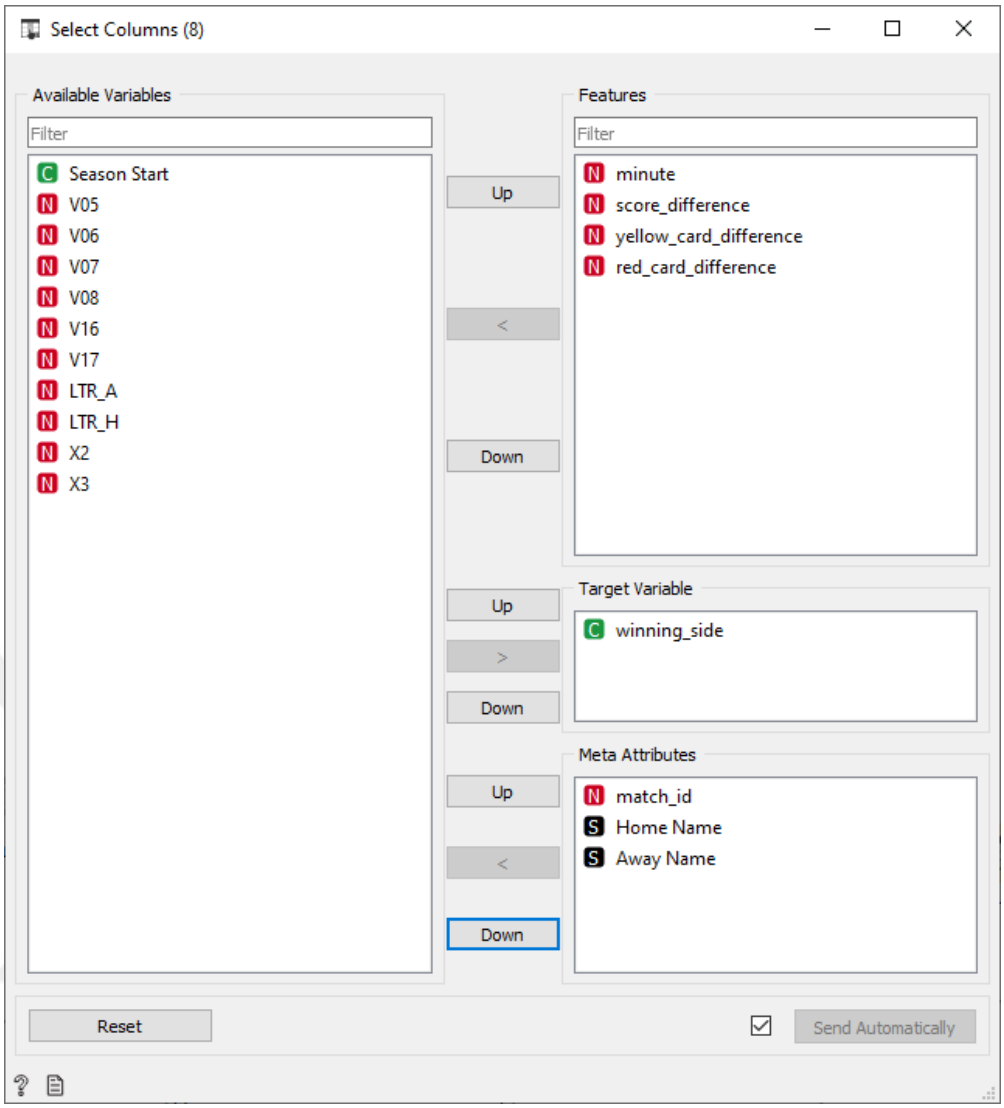
Şekil 10. 1 Maç içi model kurgusu

Aşamalar;

- Veriseti CSV formatında içeri alınmıştır. (71499 kayıt)
- Sıfırıncı yani maç önü ve 92. Dakika yani maç sonu satırlarını filtrelenmiştir.
- Eğitim seti için 2010, 2012, 2014, 2016, 2018 filtre kullanılmıştır.
- Test seti için 2011, 2013, 2015, 2019 filtre kullanılmıştır.

10.2.4 Seçilen Parametreler

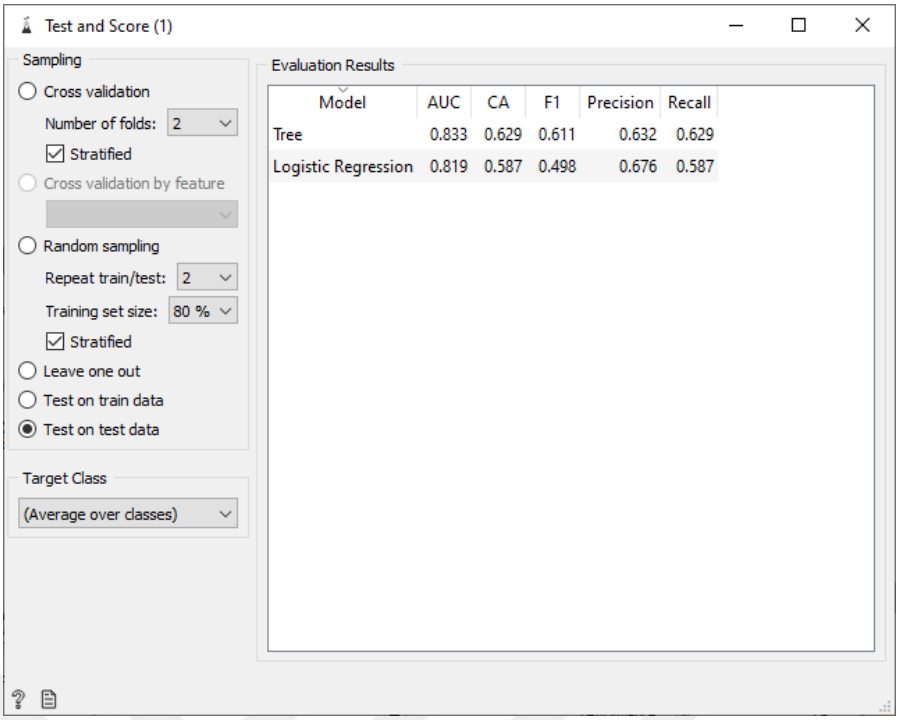
Modelin geliştirilmesi için Şekil 10.2'deki parametreler seçilmiştir:



Şekil 10. 2 Maç içi model parametreleri

10.2.5 Performans

Karar Verme Ağacı ve Lojistik Regresyon algoritmaları ile elde edilen sonuç Şekil 10.3'deki gibidir.



Şekil 10. 3 Canlı skor modeli sonuçları

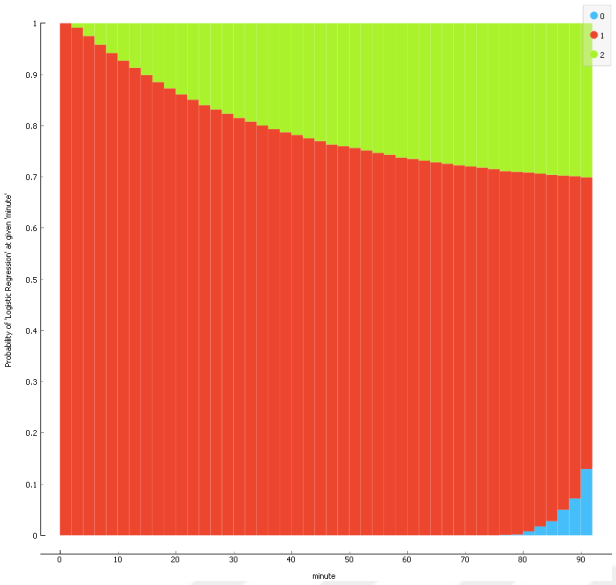
Karar verme ağacı modelinin tahminlerinin başarı oranı %63,2, gerçekleşen sonucu bilme oranı %62,9 olarak hesaplanmıştır.

Lojistik regresyon modelinin tahminlerinin başarı oranı %67,6, gerçekleşen sonucu bilme oranı %58,7 olarak hesaplanmıştır.

10.3 Dakika Bazında Tahminleme Eğilimleri

10.3.1 Lojistik Regresyon

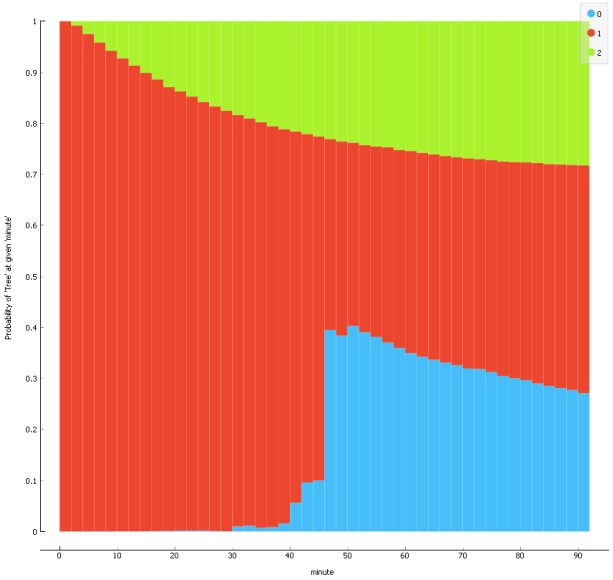
Lojistik Regresyon dakika bazında maç sonucu tahminleri Şekil 10.4'deki gibidir. Mavi alan beraberlik tahminini, kırmızı alan ev sahibi galibiyeti tahminini, yeşil alan ise deplasman takımı galibiyeti tahminini ifade etmektedir. Ancak, 80. dakikadan sonra beraberlik tahmini verilebilmiştir.



Şekil 10. 4 LR dakika bazında tahminleme eğilimleri

10.3.2 Karar Verme Ağacı

Karar Verme Ağacı dakika bazında maç sonucu tahminleri Şekil 10.5'deki gibidir. Mavi alan beraberlik tahminini, kırmızı alan ev sahibi galibiyeti tahminini, yeşil alan ise deplasman takımı galibiyeti tahminini ifade etmektedir.



Şekil 10. 5 Karar verme ağacı dakika bazında tahminleme eğilimleri

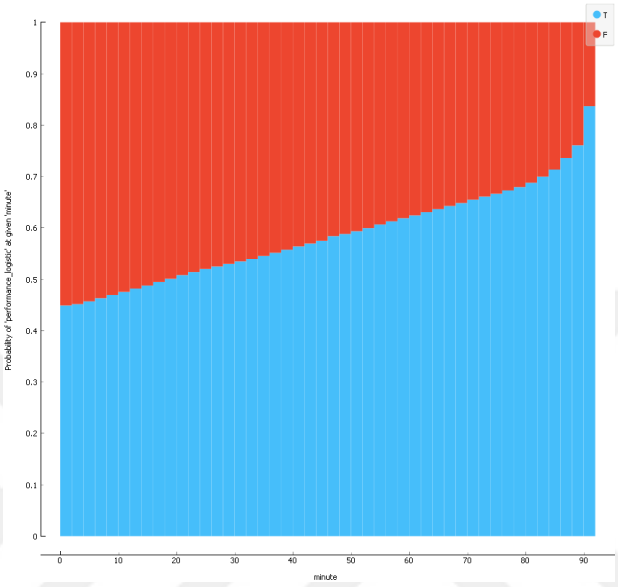
Karar Verme Ağacının 30. Dakikalara kadar beraberlik tahmini yapamadığı gözlenmiştir.

Lojistik Regresyon da ancak 80. Dakikadan sonra beraberlik tahminini çok kısıtlı olarak yapabilmiştir.

10.4 Dakika Bazında Tahminleme Performansları

10.4.1 Lojistik Regresyon

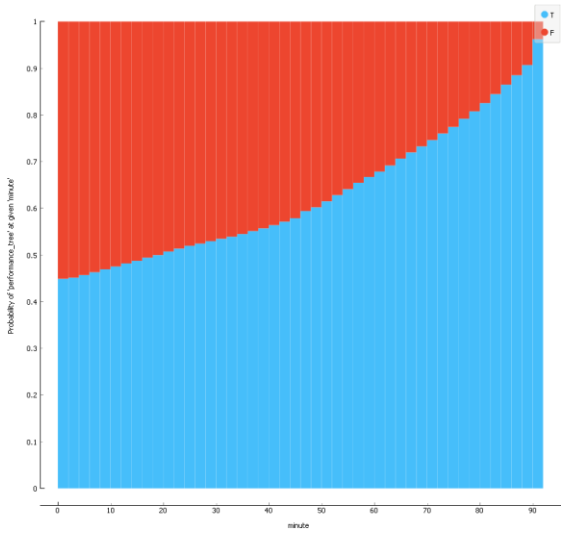
Lojistik Regresyon dakika bazında maç sonucu tahminleri Şekil 10.6'daki gibidir. Yatay eksen dakikayı, mavi alan doğru tahmin edilen alanı, kırmızı alan yanlış tahmin edilen alanı ifade etmektedir.



Şekil 10. 6 Lr dakika bazında tahminleme performansları

10.4.2 Karar Verme Ağacı

Karar Verme Ağacı dakika bazında maç sonucu tahminleri Şekil 10.7'deki gibidir. Yatay eksen dakikayı, mavi alan doğru tahmin edilen alanı, kırmızı alan yanlış tahmin edilen alanı ifade etmektedir.



Şekil 10. 7 Karar verme ağacı dakika bazında tahminleme performansları

Sadece maç içi verileri kullanılarak kurgulanan yapıda, Karar Verme Ağacı algoritması ile ancak 30. Dakika itibarı ile beraberlik tahminlerinin üretilebildiğini gözlemledik. Bu durumda modelin ilk 30 dakikasındaki tahmin etme eğilimini düzeltebilmek için maç sonu verilerinden üretilmiş parametreleri da kullanarak yeni modeller üretilmelidir.

11
CANLI SKOR MODELİ (4 ARALIKLI AYRIŞTIRMA)

Sadece maç içi parametreleri ile üretilmiş modelde, takımların ile ilgili herhangi bir bilgi bulunmamaktadır. Bu yüzden elimizdeki verisetinin makine öğrenmesi kapsamında daha iyi ayrıştırılması için, maş sonu parametrelerinin algoritmaya eklenmesi gerekmektedir.

Buradaki en büyük handikap, elimizdeki maç sayısının sınırlı olmasıdır. Eğer feaure'lar için sürekli (continous) değerler alınır ve makine öğrenmesi bu şekilde sağlanırsa, overfit durumuna düşülecektir. Bu yüzden sürekli değişkenlerin belirli sayıda aralık kullanılarak ayrıklaştırılması düşünülmüştür. Bunun sayesinde elimizdeki sınırlı sayıdaki maç verisi makine öğrenmesi için yeterli gelebilecektir.

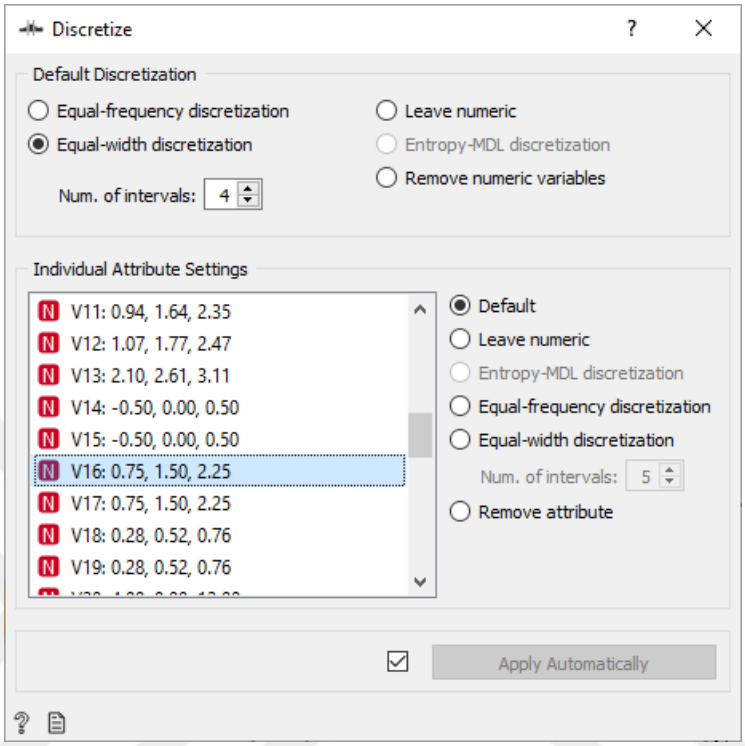
Örneğin, V16 adlı maç sonu parametresi ev sahibi takımın uzun dönem performansını göstermektedir. Eldeki verilerde yapılan analizde bu parametrenin değerinin 0,4 ile 3 arasında değişmekte olduğudur.

Bu parametresi 4 aralığa göre ayrıklaştırırsak yani 0-0,75, 0,75-1,5, 1,5-2,25 ve 2,25 ve üstü şeklinde 4 aralığa indirgersek bu durumda, 2,25 ve üstü olan takımlar aynı seviyede algılanacaktır, böylece eldeki veri miktarı artacaktır.

Eğer sürekli değişken kullanılsaydı bu durumda ondalık değerleri üzerinden Karar Verme Ağacı bir çok kural tanımlanması gerekecekti ki bu da verisetinin yetersiz veya bazı maç değerleri için overfit olmasına sebep olacaktı. Burada Karar Verme Ağacının derinlik hiperparametresi ile bir ayar yapılabilse de bu da daha generic bir tahmin yapılmasına sonuç olarak da tahminlerin kesinliğinin azalmasına sebep olacaktır.

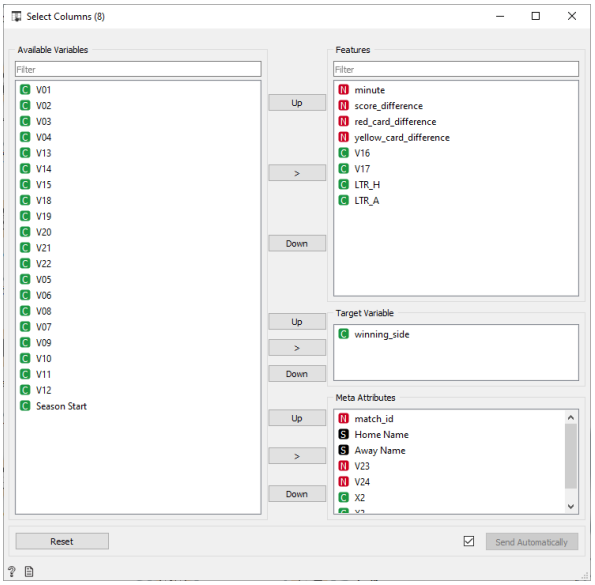
Bu kapsamda korelasyonu yüksek maç sonu parametrelerinin ayrıklaştırılıp modele eklenmesi için, tasarım şeması korunmuş sadece maç sonucu modellerinin ayrıklaştırılması için bir fonksiyon eklenmiştir.

Önce 4 aralıklı ayrıklaştırılma yapılmıştır.



Şekil 11. 1 4 aralıklı ayrıştırma

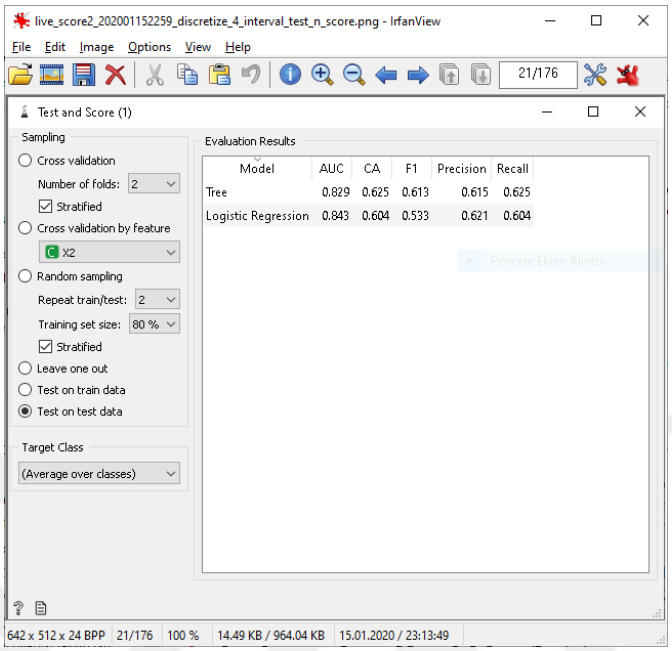
Filtreler, eğitim ve test veriseti kriterleri aynen korunmuştur. Modelin geliştirilmesi için Şekil 11.2'deki parametreler seçilmiştir:



Şekil 11. 2 4 aralıklı ayrıştırma parametre seçimi

Performans

Lojistik regresyon ve karar verme ağacı modelleri başarı oranları Şekil 11.3'de verilmiştir.



Şekil 11. 3 4 aralıklı ayrıştırma sonuçları

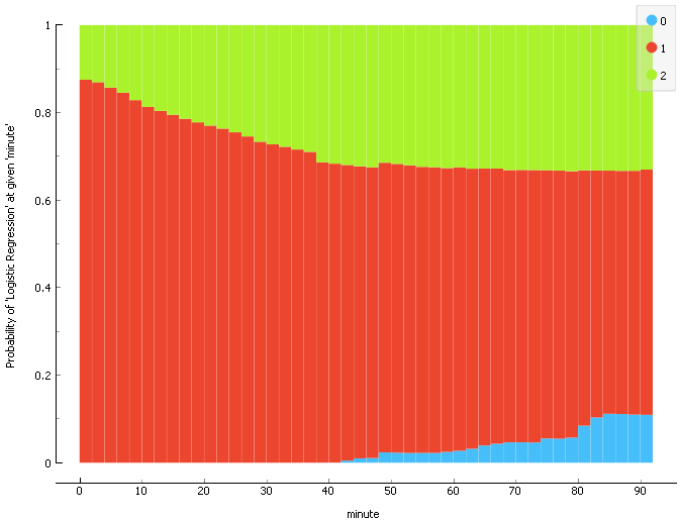
Karar verme ağacı modelinin tahminlerinin başarı oranı %61,5, gerçekleşen sonucu bilme oranı %62,5 olarak hesaplanmıştır.

Lojistik regresyon modelinin tahminlerinin başarı oranı %62,1, gerçekleşen sonucu bilme oranı %60,4 olarak hesaplanmıştır.

11.1 Dakika Bazında Tahminleme Eğilimleri

11.1.1 Lojistik Regresyon

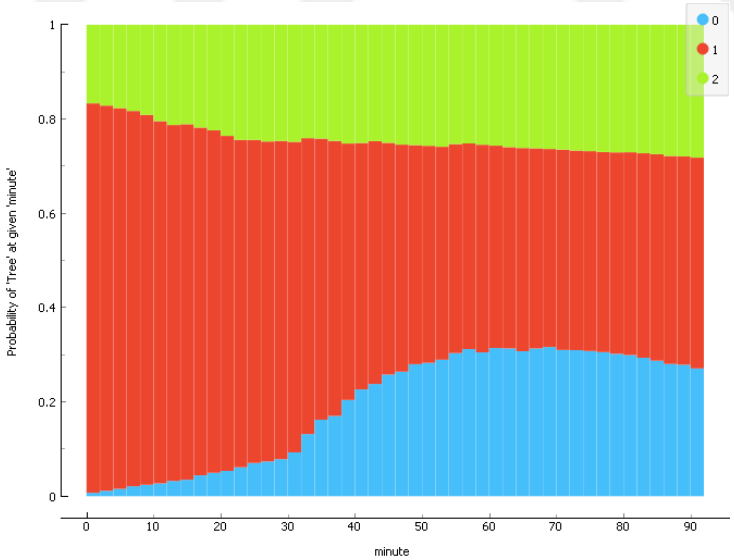
Lojistik regresyon dakika bazında maç sonucu tahminleri Şekil 11.4'deki gibidir. Mavi alan beraberlik tahminini, kırmızı alan ev sahibi galibiyeti tahminini, yeşil alan ise deplasman takımı galibiyeti tahminini ifade etmektedir. 40. dakikadan itibaren beraberlik tahmini verilebilmiştir.



Şekil 11. 4 4 aralıklı ayrıştırma lr tahminleme eğilimleri

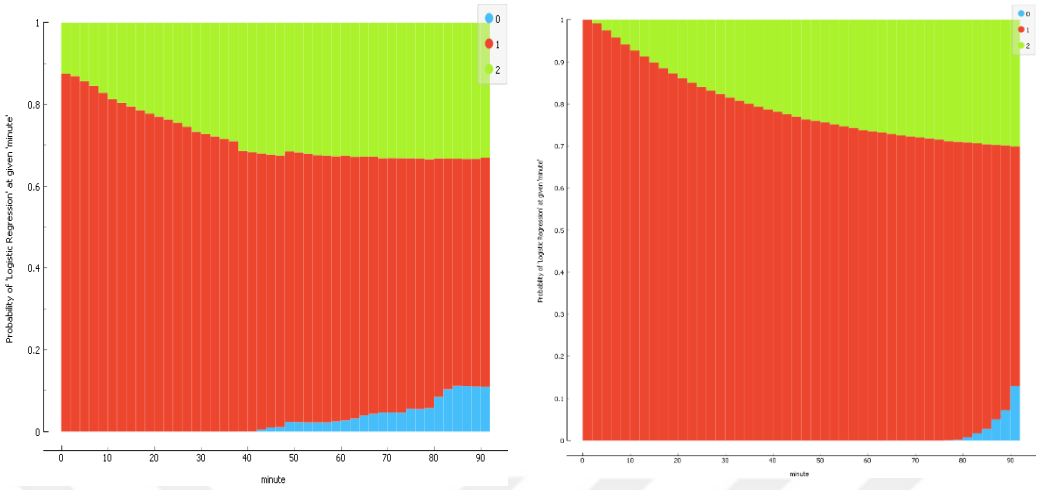
11.1.2 Karar Verme Ağacı

Karar Verme Ağacı dakika bazında maç sonucu tahminleri Şekil 11.5'deki gibidir. Mavi alan beraberlik tahminini, kırmızı alan ev sahibi galibiyeti tahminini, yeşil alan ise deplasman takımı galibiyeti tahminini ifade etmektedir. Maçın başından itibaren beraberlik tahmini verilebilmiştir.



Şekil 11. 5 4 aralıklı ayrıştırma karar verme ağacı tahminleme eğilimleri

Lojistik Regresyon için berabelik tahmin edebilme daha erken dakikalara çekilmiştir.

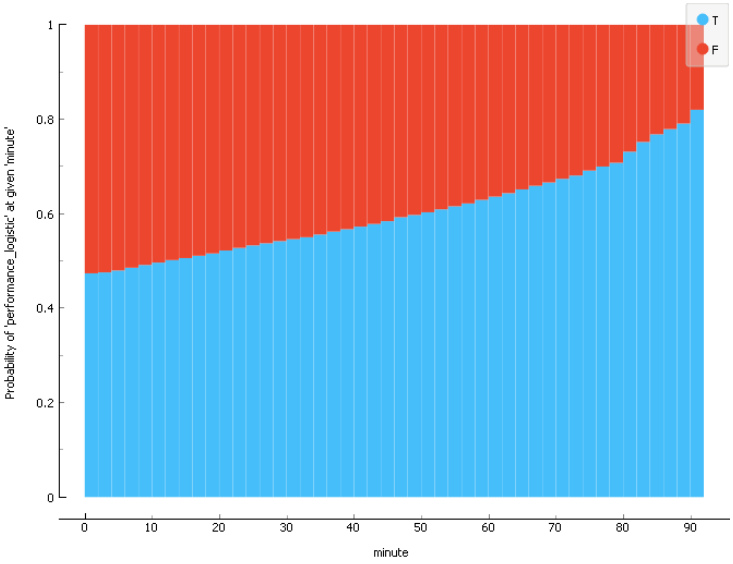


Şekil 11. 6 4 aralıklı ayrıştırma tahminleme eğilimleri karşılaştırması

11.2 Dakika Bazında Tahminleme Performansları

11.2.1 Lojistik Regresyon

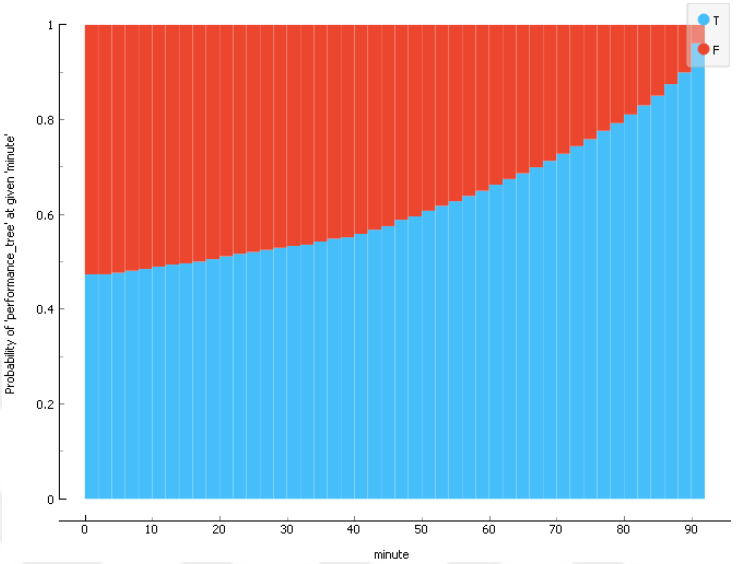
Lojistik regresyon dakika bazında maç sonucu tahminleri Şekil 11.7'deki gibidir. Yatay eksen dakikayı, mavi alan doğru tahmin edilen alanı, kırmızı alan yanlış tahmin edilen alanı ifade etmektedir.



Şekil 11. 7 4 aralıklı ayrıştırma lr tahminleme performansları

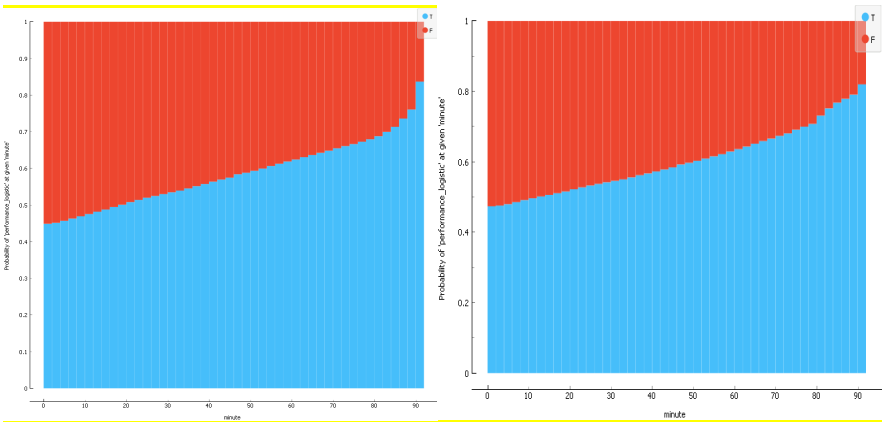
11.2.2 Karar Verme Ağacı

Karar verme ağacı dakika bazında maç sonucu tahminleri Şekil 11.8'deki gibidir. Yatay eksen dakikayı, mavi alan doğru tahmin edilen alanı, kırmızı alan yanlış tahmin edilen alanı ifade etmektedir.



Şekil 11. 8 4 aralıklı ayrıştırma karar verme ağacı tahminleme performansları

Lojistik Regresyon için berabelik tahmin edebilme daha erken dakikalara iyileşmiştir.



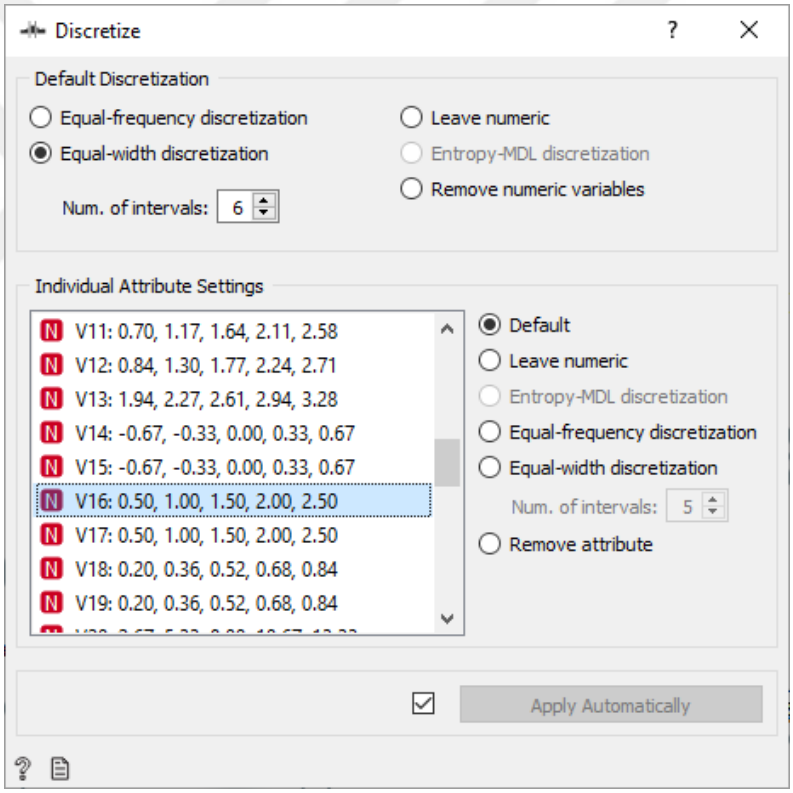
Şekil 11. 9 4 aralıklı ayrıştırma lr tahminleme performansları karşılaştırması

Karar Verme Ağacında da performansın daha erken dakikalarda iyileştiği gözlenmiştir.

12
CANLI SKOR MODELİ (6 ARALIKLI AYRIŞTIRMA)

4 aralıklı ayrıştırmadan sonra, dakika bazında daha yüksek tahmin başarı oranlarına ulaşabilme amacıyla 6 aralıklı ayrıştırma denenmiştir. Bu yöntemle de sırasıyla genel performans değerleri ölçülmüş, dakika bazında tahminleme eğilimleri ve performansları incelenmiştir.

Ayrıştırma işlemi Şekil 12.1'deki gibi Python Orange yazılımı üzerinden gerçekleştirilmiştir.

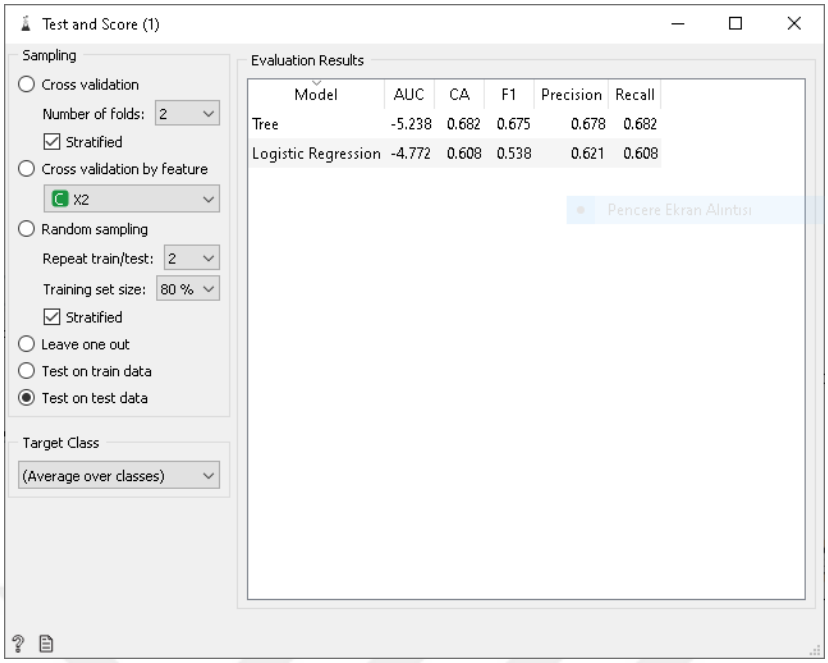


Şekil 12. 1 6 aralıklı ayrıştırma

Filtreler, eğitim ve test veriseti kriterleri aynen korunmuştur. Modelin geliştirilmesi için Şekil 12.1'deki parametreler seçilmiştir:

Performans

Karar Verme Ağacı ve Lojistik Regresyon algoritmaları ile elde edilen sonuç Şekil 12.2'deki gibidir.



Şekil 12. 2 6 aralıklı ayrıştırma sonuçları

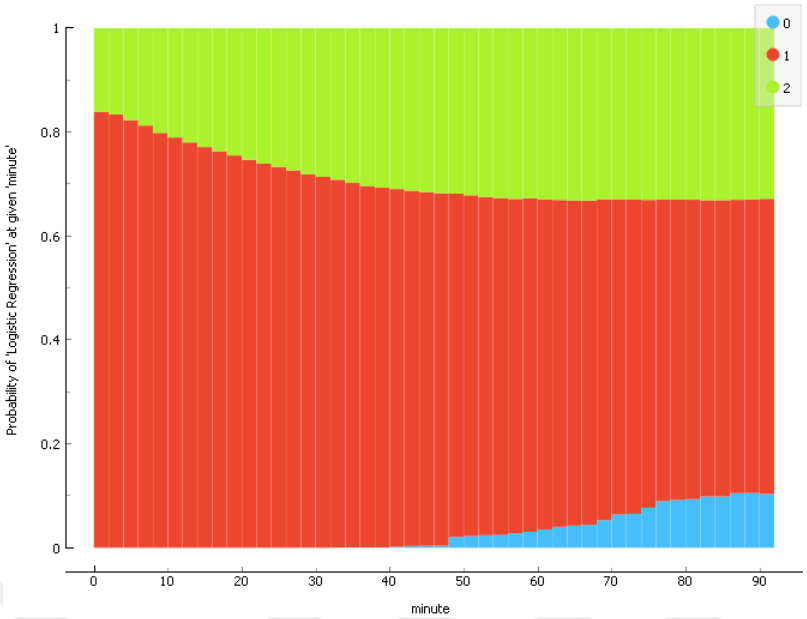
Karar verme ağacı modelinin tahminlerinin başarı oranı %67,8, gerçekleşen sonucu bilme oranı %68,2 olarak hesaplanmıştır.

Lojistik regresyon modelinin tahminlerinin başarı oranı %62,1, gerçekleşen sonucu bilme oranı %60,8 olarak hesaplanmıştır.

12.1 Dakika Bazında Tahminleme Eğilimleri

12.1.1 Lojistik Regresyon

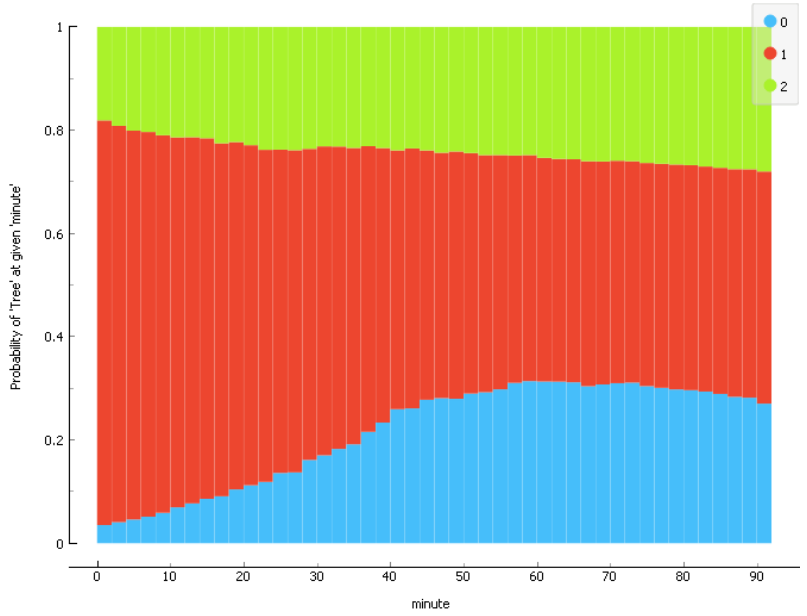
Lojistik regresyon dakika bazında maç sonucu tahminleri Şekil 12.3'deki gibidir. Mavi alan beraberlik tahminini, kırmızı alan ev sahibi galibiyeti tahminini, yeşil alan ise deplasman takımı galibiyeti tahminini ifade etmektedir. 2. yarıdan itibaren beraberlik tahmini verilebilmiştir.



Şekil 12. 3 6 aralıklı ayrıştırma lr tahminleme eğilimleri

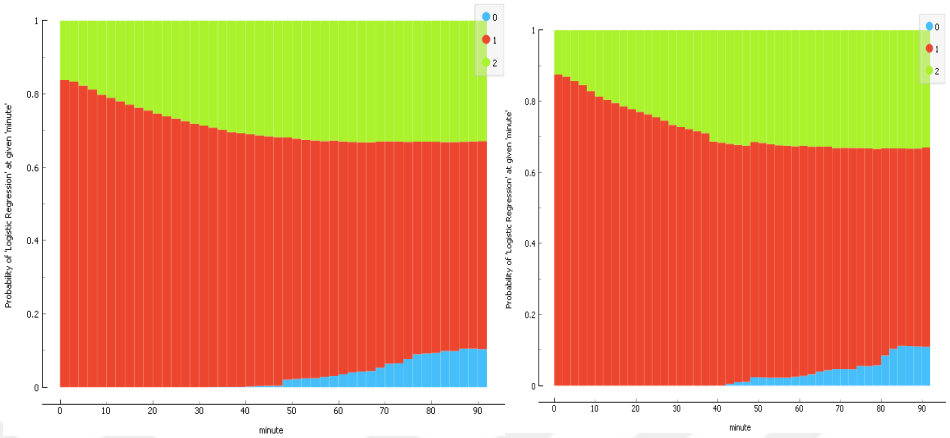
12.1.2 Karar Verme Ağacı

Karar verme ağacı dakika bazında maç sonucu tahminleri Şekil 12.4'deki gibidir. Mavi alan beraberlik tahminini, kırmızı alan ev sahibi galibiyeti tahminini, yeşil alan ise deplasman takımı galibiyeti tahminini ifade etmektedir. Maç başından itibaren beraberlik tahmini verilebilmiştir.



Şekil 12. 4 6 aralıklı ayrıştırma karar verme ağacı tahminleme eğilimleri

Lojistik Regresyon için bu işlem berabelik tahmin edebilme daha erken dakikalara istenilen ölçüde çekememiştir.

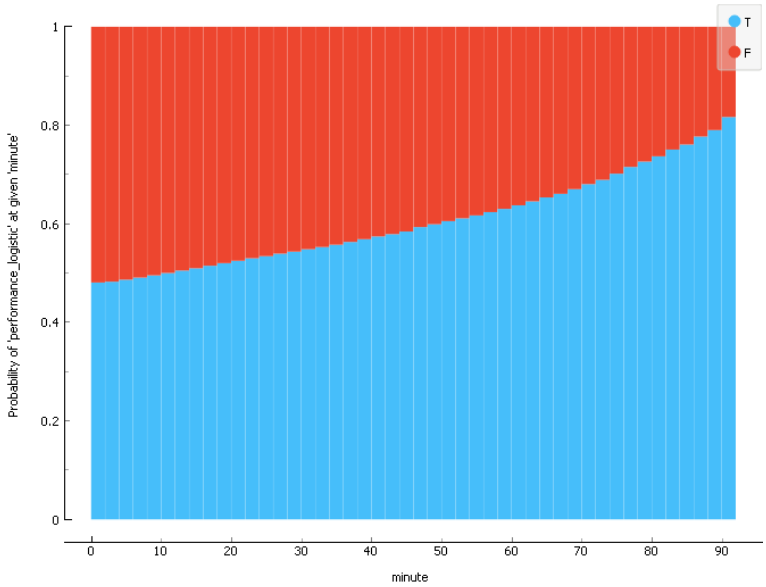


Şekil 12. 5 6 aralıklı ayrıştırma tahminleme eğilimleri karşılaştırması

12.2 Dakika Bazında Tahminleme Performansları

12.2.1 Lojistik Regresyon

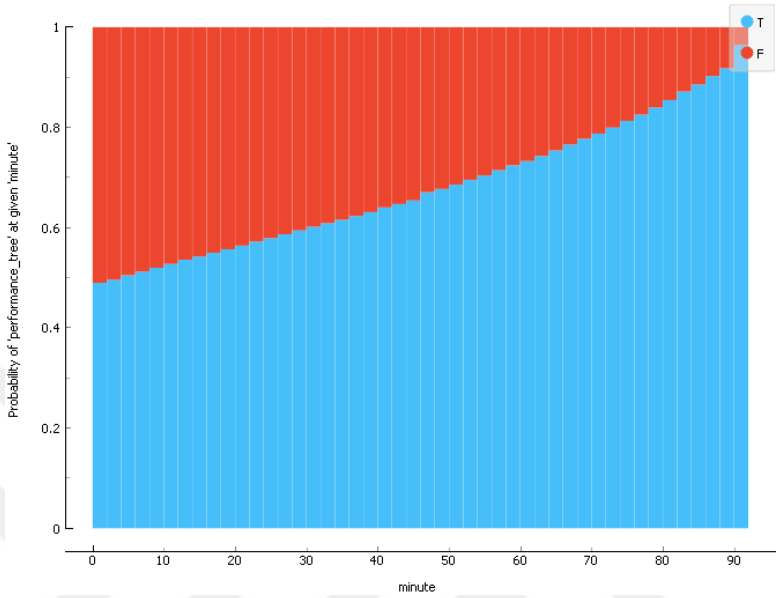
Lojistik regresyon dakika bazında maç sonucu tahminleri Şekil 12.6'daki gibidir. Yatay eksen dakikayı, mavi alan doğru tahmin edilen alanı, kırmızı alan yanlış tahmin edilen alanı ifade etmektedir.



Şekil 12. 6 6 aralıklı ayrıştırma lr tahminleme performansları

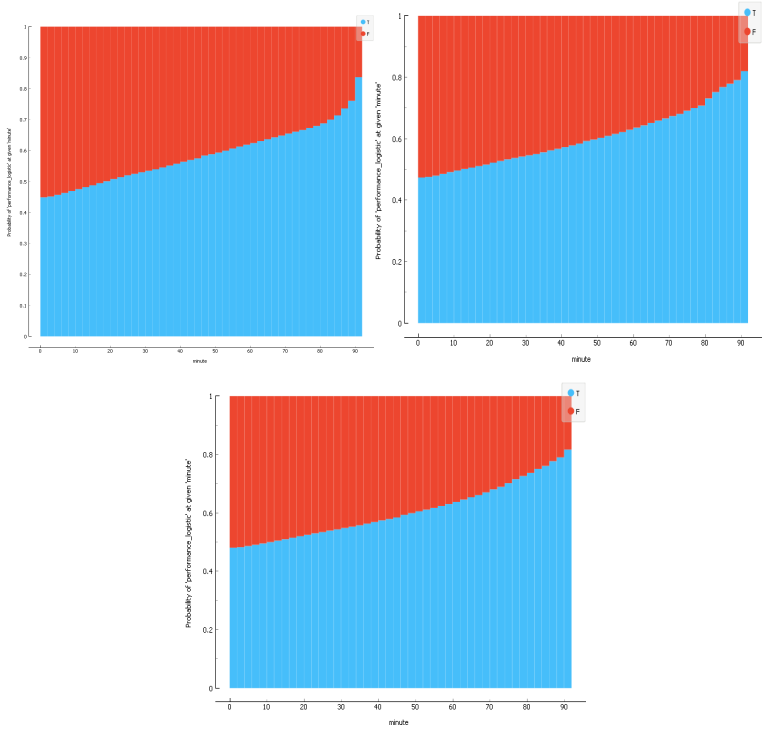
12.2.2 Karar Verme Ağacı

Karar verme ağacı dakika bazında maç sonucu tahminleri Şekil 12.7'deki gibidir. Yatay eksen dakikayı, mavi alan doğru tahmin edilen alanı, kırmızı alan yanlış tahmin edilen alanı ifade etmektedir.



Şekil 12. 7 6 aralıklı ayrıştırma karar verme ağacı tahminleme performansları

Lojistik Regresyon'da performans da erken dakikalarda beklendiği kadar iyileşmemiştir.

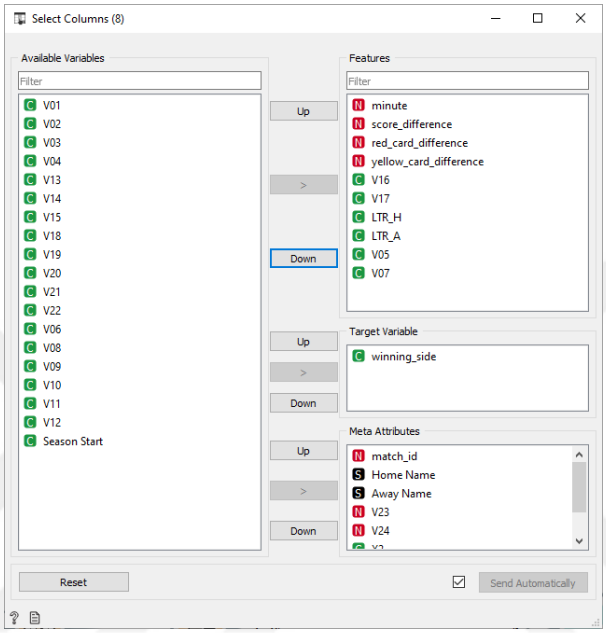


Şekil 12. 8 6 aralıklı ayrıştırma lr tahminleme performansları karşılaştırması

Karar verme ağacında da performansın daha erken dakikalarda iyileştiği gözlenmiştir.

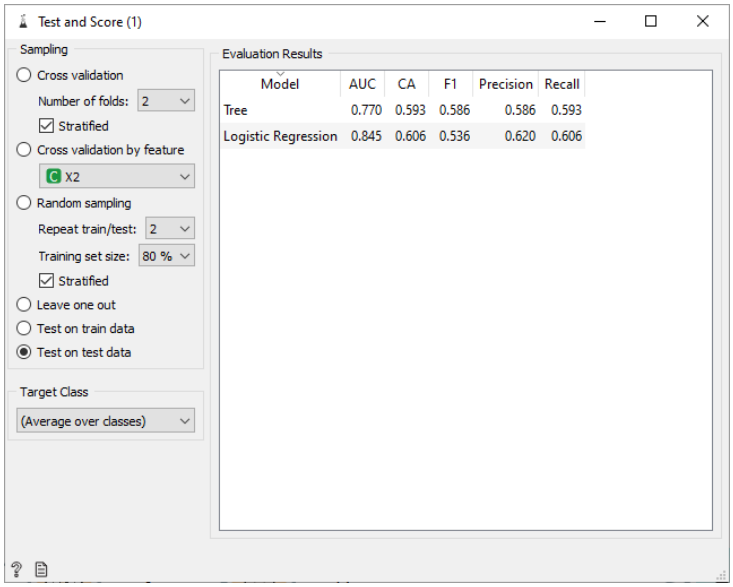
13
CANLI SKOR MODELİ (MAÇ SONU PARAMETRE SAYISINI ARTTIRMA)

Diğer tüm kurguyu değiştirmeden sadece Şekil 13.1'deki şekilde parametre sayısı arttırılmıştır.



Şekil 13. 1 Canlı skor parametre sayısını arttırma

Dakika, skor farkı, sarı ve kırmızı kart farklarına ek olarak ilk modeldeki parametreler de eklenmiştir.



Şekil 13. 2 Parametre sayısını arttırma sonuçları

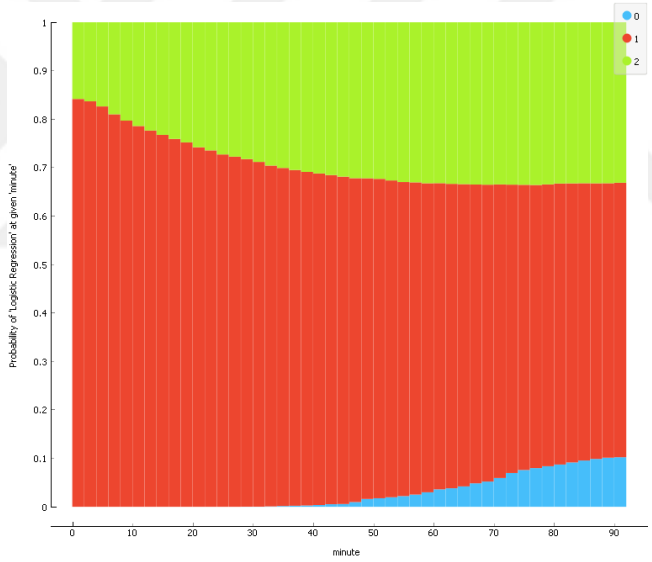
Karar verme ağacı modelinin tahminlerinin başarı oranı %58,6, gerçekleşen sonucu bilme oranı %59,3 olarak hesaplanmıştır.

Lojistik regresyon modelinin tahminlerinin başarı oranı %62, gerçekleşen sonucu bilme oranı %60,6 olarak hesaplanmıştır.

13.1 Dakika Bazında Tahminleme Eğilimleri

13.1.1 Lojistik Regresyon

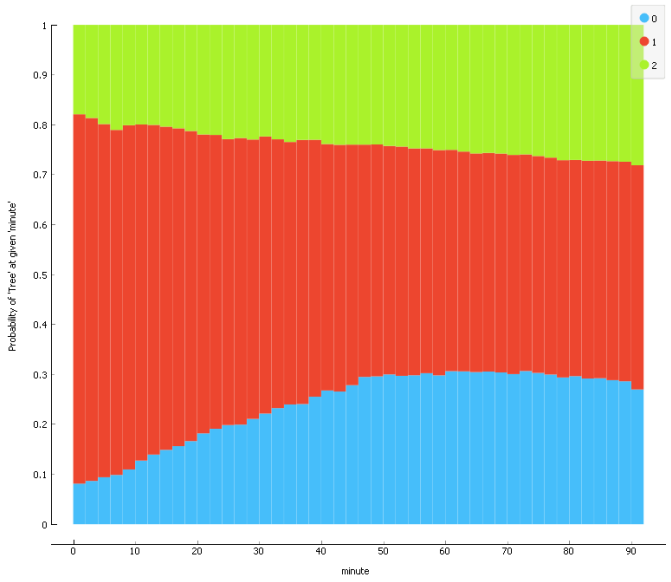
Lojistik regresyon dakika bazında maç sonucu tahminleri Şekil 13.3'deki gibidir. Mavi alan beraberlik tahminini, kırmızı alan ev sahibi galibiyeti tahminini, yeşil alan ise deplasman takımı galibiyeti tahminini ifade etmektedir. 40. dakikadan itibaren beraberlik tahmini verilebilmiştir.



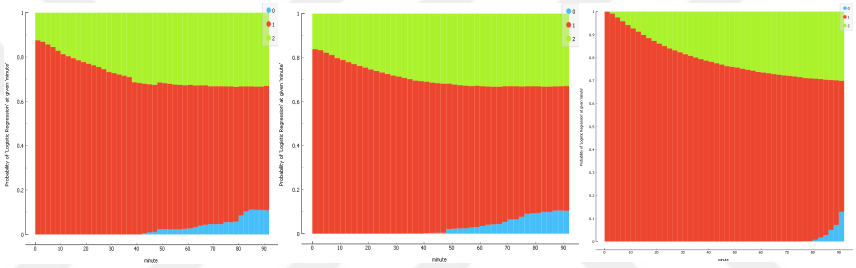
Şekil 13. 3 Parametre sayısını arttırma lr tahminleme eğilimleri

13.1.2 Karar Verme Ağacı

Karar verme ağacı dakika bazında maç sonucu tahminleri Şekil 13.4'deki gibidir. Mavi alan beraberlik tahminini, kırmızı alan ev sahibi galibiyeti tahminini, yeşil alan ise deplasman takımı galibiyeti tahminini ifade etmektedir. Maç başından itibaren beraberlik tahmini verilebilmiştir.



Şekil 13. 4 Parametre sayısını arttırma karar verme ağacı tahminlemeleri

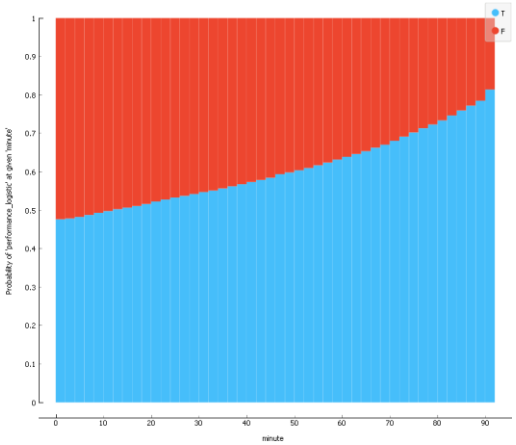


Şekil 13. 5 Parametre sayısını arttırma tahminleme eğilimleri karşılaştırması

13.2 Dakika Bazında Tahminleme Performansları

13.2.1 Lojistik Regresyon

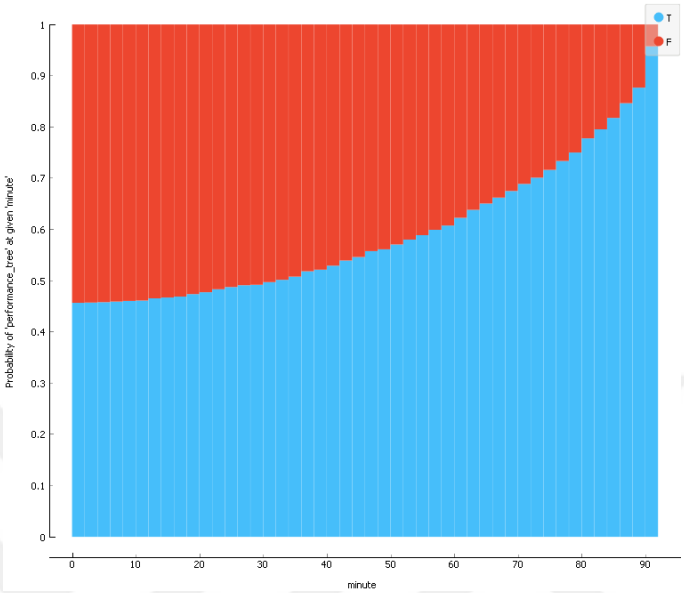
Lojistik regresyon dakika bazında maç sonucu tahminleri Şekil 13.6'daki gibidir. Yatay eksen dakikayı, mavi alan doğru tahmin edilen alanı, kırmızı alan yanlış tahmin edilen alanı ifade etmektedir.



Şekil 13. 6 Parametre sayısını arttırma lr tahminleme performansları

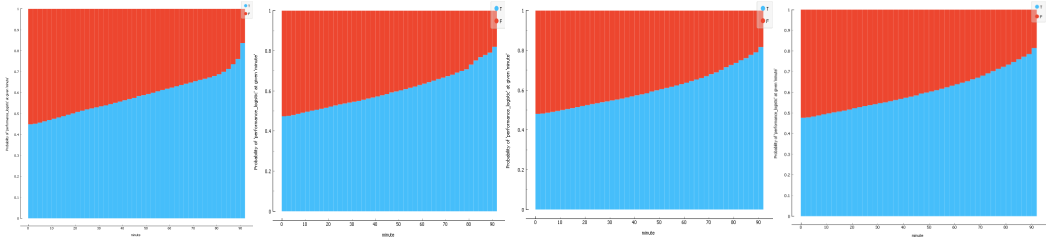
13.2.2 Karar Verme Ağacı

Karar verme ağacı dakika bazında maç sonucu tahminleri Şekil 13.7'deki gibidir. Yatay eksen dakikayı, mavi alan doğru tahmin edilen alanı, kırmızı alan yanlış tahmin edilen alanı ifade etmektedir.



Şekil 13. 7 Parametre sayısını arttırma kva tahminleme performanslar

Lojistik Regresyon'da performans da erken dakikalarda beklendiği kadar iyileşmemiştir.



Şekil 13. 8 Parametre sayısını arttırma lr tahminleme performansları

Maç sonu parametre sayısının arttırılması maç içi parametrelerinin etkisini azaltmaya başlamıştır. Karar noktası bir önceki 4 adet maç sonu parametre kullanılması olmalıdır.

14
BULGULAR VE TARTIŞMA

Maç öncesi parametreler kullanılarak maç sonucu tahminlemesi için yapılan deneyler Tablo 14.1'de gösterilmiştir. Tabloda da görüleceği üzere ilk kurgulanan model ile alınan sonuçlar yeterli görülmemiş, tahmin başarı oranını arttırmak amacıyla sırasıyla skorlama çalışması ve korelasyon analizi yapılarak parametrelerde iyileştirmeye gidilmiştir. Sonrasında gürültü azaltma çalışmaları yapılmış ve sonuçta iyileşme sağlanmıştır. Sonraki aşamada yeni parametreler denenmiştir. Öncelikle takımın 10 sezon boyunca maç başına aldığı puan ortalamasına bakılan LTR parametresi eklenmiştir. Sonrasında takımın son 3 maçını rakip takım performansını da hesaba katarak baz alan STR parametresi denenmiştir.

Tablo 14. 1 Deney takip çizelgesi

Deney	Amacı	Sonuç
İlk Model	Geçmiş çalışmaların tekrarlanıp raporlanması	Çalışmalara devam edilmelidir.
İddaa Oranları	İddaa oranlarının başarısını görmek	İddaa başarı oranları modelimiz için önemli bir referans teşkil etmektedir. Bu yüzden bu sonuçlar ilerideki model geliştirme çalışmalarında hata analizi için kullanılmalıdır.
Skorlama	Skorlama performansı	Sonuçlardan V16, V17, V18, V19 değerlerinin sonuçla olan ilişkisinin diğer parametrelere göre çok yüksek olduğu gözükmektedir. Ayrıca takip eden V05, V14, V15 ve V09 parametrelerinin de sonuç üzerinde kısmı bir etkisi olduğu gözlenmiştir. Bu parametrelerin daha etkin hale getirilmesi gerekmektedir. Korelasyon analizine geçilmelidir
Korelasyon	Parametreler Korelasyon Analizi	Sonuçlardan parametre çiftleri arasında korelasyon oranları yüksek gözlenmiştir. Bu seçilen parametrelerin birbirinden bağımsız olmadıkları ve regresyon/gruplandırma modelinde tekrarlara yol açtığı anlaşılmıştır. Bağımsız parametre setlerinin kurgulanması gerekmektedir.
Gürültü	Gürültü Azaltma	Ana verisetinin iki ayrı verisetine ayrıldıktan sonra çok daha iyi modellenebildiği fark edilmiştir. Bu kapsamda Büyük ve Küçük ligler için 2 ayrı model çalışılacak, ayrıca Neural Network (Sinir Ağları) parametreleri de optimize edilerek beraberlik için ayrı bir kurgu yapılacaktır. Ayrıca verisetindeki eksik veriler ile gerçek sıfır değerlerinin birbirinden ayrıştırılması için verisetinin düzenlenmesi gerekmektedir.
LTR	Uzun Dönem Katsayı	Bu parametre seti HomeId ve AwayId 'i doğrudan kullanmak yerine takımların kategorize edilebilmesini sağlamıştır. LTR_H ve LTR_A değerleri için olduğu kadar V09, V10, V11 ve V12 parametreleri için de tanımsız değerin -1 yapılması, verisetinin incelenmesini kolaylaştıracak ve ayrıştırma işlemine yardımcı olacaktır.

Sonraki aşamada ise maç içi veriler kullanılarak maçın 0. Dakikasından 92 dakikasına kadar tahminleme çalışması yapılmıştır. Bu aşamada ise lojistik regresyon ve karar verme ağacının daha yüksek oranlı tepki verdiği gözlenmiştir. Sonrasında ise 4 aralıklı ayrıştırma ve 6 aralıklı ayrıştırma ile deneyler tekrarlanmıştır.

İlk model sonucunda karar verme ağacının 30. Dakikadan itibaren beraberlik tahmini verebildiği, lojistik regresyon modelinin ise 80. Dakikaya kadar beraberlik tahmini veremediği gözlenmiştir. 4 aralıklı ayrıştırma sonrasında ise lojistik regresyonda 40. Dakikada, karar verme ağacında ise maçın başından itibaren beraberlik tahmini verilebildiği görülmüştür.

15
SONUÇ VE ÖNERİLER

Maç öncesi ve maç içi tahminler konusunda yapılan deneyler sonucunda özellikle parametre optimizasyonunun sonucu her seferinde iyileştirdiği görülmüştür. Maç içi parametreleri kullanmadan yani maç başlamadan önce yapılan yapılan tahminlerin maç başladıktan sonraki tahminlerden farklı olacağı aşikardır. Bu yüzden, maç önü tahminleri ile maç içi tahminlerinin bir şekilde harmanlanması gerekmektedir.

Burada tavsiye edilen, yukarıda bahsedilen maç olaylarından biri gerçekleşene kadar, Karar Verme Ağacı algoritması ile maç öncesi tahminin ağırlıklı ortalamasının alınmasıdır. Burada ağırlık katsayıları dakikalar geçtikçe maç içi modeline doğru evrilmelidir.

Toplam evrilme zamanı 30 dakika öngörülürse, bu durumda, maç içi modelin ağırlık katsayısı = (dakika -1) / 30 olur.

Maç önü tahminin ağırlık katsayısı = 1 - maç içi modelin ağırlık katsayısı olarak hesaplanacaktır.

30. dakika sonrası itibarı ile tamamen maç içi modelin tahminleri devrede olacaktır.

Eğer ilk 30 dakika içerisinde yukarıda bahsedilen maç olaylarından biri gerçekleşirse, direkt olarak maç içi modele geçilmelidir.

Bu çalışma daha geniş bir data ağı ile alanında yapılacak çalışmalar için öncü niteliğinde olacaktır. Günümüzde dünyada Eldo (Amerikan Futbolu), Whiteball Analytics (Kriket) gibi daha az popüler spor dallarına yönelik olasılık analizi sunan yabancı firmalar bulunmaktadır. Türkiye açısından ise, özellikle medya sektöründeki kurumsal müşterilerin analizlerinde Opta (Perform Group), Brodage gibi yurt dışı kaynaklı data sunucularını kullandıkları görülmektedir. Spor kulüplerinde ise, kadro alternatiflerini de içeren senaryo analizi araçlarının kullanımının sınırlı olduğu görülmektedir. İş modellerini müsabaka tahminleme üzerine kuran tipster sitelerinin ve Maçkolik, Nesine.com gibi bireysel kullanıcılara hitap eden firmaların ise detaylı istatistikli modellere dayanmayan editöryel

yorumlar ile ilerlediği görülmektedir. Bu çalışma ise bu alanda gerek veri-müsabaka sonuçları arasındaki detaylı korelasyon analizleri, gerekse makine öğrenmesi prensiplerini uygulayan dinamik öğrenme yapısı ile fark yaratacaktır.

KAYNAKÇA

[1] Bunker R.P., Thabtah F. 2017. “A Machine Learning framework for sport result prediction”, Computing and Informatics, https://doi.org/10.1016/j.aci.2017.09.005.

[2] Krauss C., Do X. A., Huck N. 2017. “Deep neural networks, gradient-boosted trees, random forests: Statistical arbitrage on the S&P 500”, European Journal of Operational Research, 259, 689-702.

[3] Gustavo E. A. P. A., Batista R. C. Prati, Monard M. C. 2004. “A Study of the Behavior of Several Methods for Balancing Machine Learning Training Data”, Sigkdd Explorations, Volume 6, Issue 1 - Page 20.

[4] Dietterich T. 1995. “Overfitting and Undercomputing in Machine Learning”, ACM computing surveys (CSUR), Vol 27, No 3.

[5] Jabbar H., Khan R.Z. 2015. “METHODS TO AVOID OVER-FITTING AND UNDER-FITTING IN SUPERVISED MAKİNE ÖĞRENMESİ (COMPARATIVE STUDY)”, Computer Science, Communication & Instrumentation Devices.

[6] Dreiseitla S., Machadob L.O. 2002. “Lojistik Regresyon and Artificial Neural Network Classification models: a methodology review”, Journal of Biomedical Informatics, Volume 35, Issues 5–6, Pages 352-359.

[7] Stekler H. O., Sendor D., Verlander R. 2010. “Issues in sports forecasting”, International Journal of Forecasting, 26, 606-621.

[8] Strumbelj E., Vračar P. 2012. “Simulating a basketball match with a homogeneous Markov model and forecasting the outcome”, International Journal of Forecasting, 28, 532-542.

[9] Joseph A., Fenton N.E., Neil M. 2006. “Predicting football results using Bayesian nets and other Machine Learning techniques”, Knowledge-Based Systems, 19, 544-553.

[10] E. USTAOĞLU, "Hedonic Price Analysis of Office Rents: A Case Study of the Office Market in Ankara”, Yüksek Lisans Tezi, Orta Doğu Teknik Üniversitesi, Sosyal Bilimler, Ankara, 2003.

[11] Maszczyk A., Gołaś A., Pietraszewski P., Roczniok R., Zając A., Stanula A. 2014. “Application of Neural and Regression Models in Sports Results Prediction”, Procedia - Social and Behavioral Sciences, 117, 482-487.

[12] Karaoğlu B. 2015. “Makine Öğrenmesi ile Spor Karşılaşmalarının Modellenmesi”, EMO Bilimsel Dergi, 5, 9.

[13] Park Y. J., Kim H. S., Kim D., Lee H., Kim S. B., Kang P. 2017. “A deep learning-based sports player evaluation model based on game statistics and news articles”, Knowledge-Based Systems, 138, 15-26.

TEZDEN ÜRETİLMİŞ YAYINLAR

İletişim Bilgisi: ozanbarlas91@gmail.com

Konferans Bildirileri

1. Barlas O., Demirel T. 2020. "Makine Öğrenmesi Yöntemleri Kullanılarak Futbol Müsabakalarının Analiz Edilmesi", International Marmara Sciences Congress (IMASCON 2020 Spring), 280-286.