




## Türkiye ve İtalya Voleybol Süper Ligleri 2013-2020 İstatistik Verilerinin Veri Madencilięi Yöntemleriyle Analizi\*

### Analysis of Turkey and Italy Volleyball Super Leagues 2013-2020 Statistical Data with Data Mining Methods\*

Emre KOMAR\*\*   
Erol EĞRİOĞLU\*\*\*   
Kıvanç SEMİZ\*\*\*\* 

#### Öz

Bu arařtırmanın amacı Türkiye ve İtalya'da 2013-2020 yılları arasında oynanan erkek ve kadın voleybol süper lig maçlarının veri madencilięi yöntemleri ve yapay sinir aęları modelleriyle maç sonu tahminleri ve maç sonucuna etki eden deęişkenlerin incelenmesidir. Bu arařtırma korelasyonel tarama arařtırmasıdır. 1144 tanesi Türkiye erkek, 1142 tanesi Türkiye kadın, 1122 tanesi İtalya erkek, 1066 tanesi İtalya kadın olmak üzere toplamda 4474 müsabaka veri madencilięi yöntemiyle Matlab'da bulunan "Statistics and Machine Learning Toolbox", "Neural Network Pattern Recognition Toolbox" araç paketleri ve "Classification Learner" uygulamasından yararlanılarak analiz edilmiştir. Sınıflama yöntemleri, tüm maçlar için, cinsiyete ve ülkelere göre farklı uygulanmış ve "20 fold cross validation (20 kat çapraz doęrnlama)" teknięi ile her bir maça kendi içerisinde etki eden 10 farklı deęişken ile sınıflama doęrulukları belirlenmiştir. Arařtırma sonucunda kullanılan teknikler ile maç sonuçlarını doęru tahmin edebilme oranının ortalama %90 olduęu saptanmıştır. Müsabaka sonucuna en çok etki eden deęişkenlerin deplasman takımlarının oyun içi varyasyonları olduęu görülmüştür. Elde edilen bulgulara göre; veri madencilięi yöntemleri olarak destek vektör makineleri en yakın k-komşu, karar aęaçları, ensemble yöntemler gibi bazı sınıflandırıcılar ve çok katmanlı algılayıcı yapay sinir aęı modellemesiyle yapılan müsabaka sonucu doęru tahmin oranlarının %90'a yakın olduęu görülmüştür. Erkek maçlarına bakıldığında maç sonucuna en çok etki eden deęişkenler

\* Bu çalışma 2023 yılında Giresun Üniversitesi Sağlık Bilimleri Enstitüsü'nce Emre Komar tarafından "Türkiye ve İtalya Voleybol Süper Ligleri 2013-2020 İstatistik Verilerinin Veri Madencilięi Yöntemleri ile Analizi" adlı tez çalışmasından üretilmiştir.

\*\* Yüksek lisans öğrencisi, Giresun Üniversitesi, Sağlık Bilimleri Enstitüsü, Beden Eğitimi ve Spor ABD, Giresun, Türkiye, emre.komar3@gmail.com, ORCID: 0000-0002-6037-5344,

\*\*\* Prof. Dr., Giresun Üniversitesi, Fen Edebiyat Fakültesi, İstatistik Bölümü, Giresun, Türkiye, erol.egrioglu@giresun.edu.tr, ORCID: 0000-0003-4301-4149

\*\*\*\* Dr. Öğ. Üyesi, Giresun Üniversitesi Spor Bilimleri Fakültesi, Antrenörlük Bölümü, Giresun, Türkiye, kivanc.semiz@giresun.edu.tr, ORCID: 0000-0003-3051-4814

deplasman takımlarının blokları, hücumları ile servis sayıları olduđu görölmektedir. Kadın maçlarına bakıldığında ise maçların sonucuna etki eden en önemli iki faktör sırasıyla deplasman takım hücumları ve deplasman takım servis sayıları olduđu görölmektedir.

**Anahtar Kelimeler:** Veri Madenciliđi, Voleybol, Büyük Veri.

### Abstract

The purpose of this research was to calculate the percentages of the variables affecting the end of the match predictions and the match result with data mining methods and artificial neural network models of the men's and women's volleyball super league matches played in Turkey and Italy between the years 2013-2020. This research is a correlational survey research. "Statistics and Machine Learning Toolbox" and "Neural Network Pattern Recognition Toolbox" tool packs in Matlab, which has a total of 4474 competition data mining programs, 1144 of which are male in Turkey, 1142 are female of Turkey, 1122 are male in Italy, and 1066 are female in Italy. It was analyzed using the "Classification Learner" application. Classification methods were applied differently for all matches, according to gender and country, and classification accuracies were determined with 10 different variables affecting each match within itself, using the "20-fold cross-validation" technique. As a result of the research, it was determined that the average rate of predicting the match results with the techniques used was %90. It has been seen that the variables that most affect the result of the competition are the in-game variations of the away teams. According to the obtained; as data mining methods, the correct prediction rates are close to 90% as a result of the contention made with support vector machines, nearest k-neighbor, decision tree, some classifiers such as ensemble routing and multi-source emitter neural network routing. Regarding the men's matches analysis, it is seen that the variables that most affect the match result are the blocks of the away teams, their attacks and the number of server. According to the women's matches analysis, it was seen that the two most important factor affecting the results of the matches are the away team attacks and the away team serve numbers, respectively.

**Keywords:** Data Mining, Volleyball, Big Data,

## GİRİŞ

Bilgi çađı olarak adlandırılan günümüz dünyasında hemen her konuda üretilen ve depolanan verilerin büyüklüğü giderek artan bir hızla devam etmektedir (Özekes, 2003). Teknolojik imkânların artmasıyla beraber veri depolama ve depolanan veriler üzerinden işlem yapabilme kapasitesinde çok büyük gelişmeler olmuştur. (Sevindik, Kayışlı ve Ünükahraman, 2012). Veri madenciliđi; depolanmış büyük veri yığınlarından farklı yazılım ve teknikler kullanılarak işlenmek amacıyla bilgilerin alınıp analiz edilmesi ve kümeleme, sınıflama gibi yöntemlerle anlamlı hale getirilmesi için kullanılan bir teknik olarak görölmektedir (Baykal, 2006). Bu tekniğin kullanılabilmesi için gerekli olan verinin doğru şekilde depolanmış olması gerekmektedir. Son yıllarda tüm dünyada kullanılmaya başlayan veri madenciliđi uygulamaları işletmelerin verimlilik ve rekabet gücünü arttırmakta ve daha kararlı bir şekilde iş gücü üretmelerine olanak sağlamaktadır (Akpınar, 2018). Veri Madenciliđi; veri tabanları hazırlama ve karar verme mekanizmaları oluşturma, işletmelere yönelik Pazar arařtırmaları yapma, müşteri potansiyelleri belirleme, pazarlama sektöründe müşterilerin satın alma durumlarını belirleme, bankacılık sektöründe müşterileri gruplandırma, borç ödeme, kredi kullanımı gibi geçmişe yönelik durumlarını inceleme ve yeni portföyler oluşturma, eğitim alanında öğrencilerin geriye dönük taramaları, başarı oranları, ileriye dönük hedef belirleme, sağlık alanında, hasta taramaları, yeni tedavi yöntemleri uygulama gibi birçok alanda hızlı ve güvenilir bilgiye ulaşmak için kullanılmaktadır (Baykal, 2006). Tahmin edici modellerde, sonuçları bilinen verilerden hareket

edilerek bir model geliştirilmesi ve kurulan bu modelden yararlanılarak sonuçları bilinmeyen veri kümeleri için sonuç değerlerin tahmin edilmesi amaçlanmaktadır (Zhong ve Zhou, 1999).

Spor müsabakalarının hemen hepsinde tutulan kayıtlar teknolojik gelişmeler sayesinde önemli birer veri tabanı oluşturmaları ve bu verilere ulaşmanın kolay ve hızlı olması müsabaka analizi konusunda büyük ilerleme kaydedilmesine yol açmıştır (Schumaker, Solieman ve Chen, 2010). Elde edilen bu verilerin analizleri sportif performans ve gelişim açısından büyük önem arz etmektedir (Millington ve Millington, 2015). Günümüzde spor kulüplerinde oyuncu performanslarını arttırmak, müsabaka kazanmaya yönelik hamleler yapmak gibi birçok farklı alanda teknolojik gelişmeleri kullanmaya özen gösterilmektedir (Corsten ve Roschen, 2018). Birçok spor branşında kaydedilen çok fazla miktarda veri bulunmasından dolayı spor alanında da veri madenciliği kullanımı büyük önem kazanmaktadır (Murathan ve Devecioğlu, 2018). Her ne kadar futbol ve basketbol günümüzün veri madenciliği kullanımı konusunda popüler branşları olsa da voleybolda apriori algoritması (birliktelik analizlerinin yapıp birliktelik kurallarının ortaya çıkartılması konusunda en çok bilinen ve kullanılan algoritmadır) kullanılarak teknik-taktik analizi (Li, 2018), yüzmede sinir ağları kullanılarak sensör teknolojileri geliştirilmekte (Oghi, Kanede ve Takakura, 2014), hentbolda yine sinir ağları yardımı ile aksiyon tahminleri yapılmaktadır (Hassan, Schrapf ve Tilp, 2017).

Bu araştırmada 2013-2020 yılları arasında Türkiye ve İtalya'da oynanan erkek ve kadın voleybol süper lig maçlarının veri madenciliği yapmaya da olanak tanıyan bir program olan Matlab ve yapay sinir ağları modelleriyle maç sonu tahminleri ve maç sonucuna etki eden değişkenlerin yüzdelerinin hesaplanması amaçlanmıştır. Bu amaç doğrultusunda ortaya çıkan analiz sonuçlarından yola çıkılarak takımların rakiplerinin hangi özelliklerine göre hangi antrenman metodlarına ağırlık vermeleri, hangi antrenman yöntemlerinin kullanılmasının takım başarısına etkili olacağı konusunda fikir yürütülebilmesi amaçlanmıştır.

## **YÖNTEM**

### **Araştırma Modeli**

Bu araştırma farklı değişkenlerin birbirleriyle olan ilişkilerini araştırmak amacıyla korelasyonel tarama araştırması olarak tasarlanmıştır.

### **Evren-Örnekleme**

Yapılan bu çalışmada 2013-2020 yılları arasındaki 7 sezonda Türkiye ve İtalya voleybol süper liglerinde oynanan erkek ve kadın maçlarının tamamı (yarıfinal ve finaller dâhil 4474 maç) incelenmiştir. Mevcut maçların 1144 tanesi Türkiye erkek, 1142 tanesi Türkiye kadın, 1122 tanesi İtalya erkek, 1066 tanesi İtalya kadın müsabakalarıdır.

## Veri Toplama Araçları

Müsabakaların data volley istatistik programı ile tutulan istatistik bilgileri ve analize tabi tutulan hücum, blok, servis karşılama, ace (servisten alınan direkt sayı), servis hatası gibi değişkenler Türkiye (TVF,t.y) ve İtalya (FIP,t.y) voleybol federasyonlarının resmi internet sitelerinin açık arşiv kayıtlarından temin edilmiştir.

## Verilerin Analizi

Makalenin amacı veri madenciliği alanına katkı sağlamak değildir. Veri madenciliği yöntemlerinin uygulanmasında standart prosedürler verinizin içeriğine göre değişebilmektedir. Yani her zaman veri madenciliğinin uygulanmasında standart veri önışlem tekniklerine ihtiyaç duyulmamaktadır. Bu çalışmada problemin ortaya koyulması, çalışmanın amacı ve dizaynı ve ilgili alana ait istatistikler spor bilimlerindeki arařtırmacılar tarafından gerçekleştirilirken, veri madenciliği yöntemleri alanında uzman ve makine öğrenmesi alanında birçok özgün makalesi olan yazarın desteği ile gerçekleşmiştir. Çalışmada hem spor hem de istatistik alanında uzmanlardan oluşan bir ekip yer almıştır.

Makalede ilgili spor alanındaki istatistiksel veriler değerlendirmeye tabi tutularak, Matlab araç paketlerinden yararlanmak suretiyle bazı veri madenciliği yöntemlerinin bu alandaki uygulanabilirliğini göstermektedir. Bunun yanında skora etki eden değişkenler ve bunların önem sırasıda yine bu veri madenciliği yöntemleri ile belirlenmiştir. Makalenin amacı veri madenciliği alanında yani bir yöntem geliştirmekten ziyade mevcut veri madenciliği yöntemlerinin spor alanında uygulanabilirliğinin arařtırılmasıdır.

Sonuçların analiz edilmesinde, Matlab'da bulunan "Statistics and Machine Learning Toolbox", "Neural Network Pattern Recognition Toolbox" araç paketleri ve "Classification Learner" uygulamasından yararlanılmıştır. Toplam incelenen maç sayısı 4474 olmaktadır. Uygulamada sınıflama yöntemleri;

- Tüm maçlar için
- Cinsiyete göre
- Ülkelere göre
- Hem ülke hem Cinsiyet alt gruplarında

uygulanmış olup "20 fold cross validation (veri kümesinin 20 farklı gruba ayrılır ve gruplardan biri test diğerleri eğitim seti olarak kullanılır ve model eğitilir)" tekniği ile her bir maça kendi içerisinde etki eden 10 farklı değişken ile sınıflama doğrulukları belirlenmiştir. Sınıflama doğrulukları arasında cinsiyete, ülkelere ve hem ülke hem cinsiyetlerin etkileşimine göre anlamlı fark olup olmadığı Mann Whitney U testi ve Kruskal Wallis H-Testi isimli parametrik olmayan istatistiksel hipotez testleri ile belirlenmiştir.

İlk olarak tüm veri için temel bileşenler analizi (principal component analysis, PCA) uygulanarak ve uygulanmadan tüm sınıflandırma yöntemleri uygulanmış ve doğru sınıflandırma yüzdeleri elde edilmiştir. Uygulamada kullanılan 20 adet sınıflama yöntemleri aşağıda verilmiştir. Yöntemlerin

bazıları aynı yöntem grubunda yer alabilmektedir. Yöntemler genel olarak karar ağaçları, destek vektör makinaları, en yakın k-komşu algoritmaları, ensemble yöntemlerden oluşmaktadır. Bu yöntemler literatürdeki en popüler sınıflandırma yöntemleri olduğundan tercih edilmiştir.

#### Fine Tree – Karar Ağacı Yöntemi

- Medium Tree – – Karar Ağacı Yöntemi
- Coarse Tree – – Karar Ağacı Yöntemi
- Linear SVM – Destek Vektör Makinası
- Quadratic SVM – Destek Vektör Makinası
- Cubic SVM – Destek Vektör Makinası
- Fine Gaussian SVM – Destek Vektör Makinası
- Medium Gaussian SVM – Destek Vektör Makinası
- Coarse Gaussian SVM – Destek Vektör Makinası
- Fine KNN – En Yakın K-Komşu Algoritması
- Medium KNN – En Yakın K-Komşu Algoritması
- Coarse KNN – En Yakın K-Komşu Algoritması
- Cosine KNN – En Yakın K-Komşu Algoritması
- Cubic KNN – En Yakın K-Komşu Algoritması
- Weighted KNN – En Yakın K-Komşu Algoritması
- Ensemble Boosted Trees – Ensemble Yöntem
- Ensemble Bagged Trees – Ensemble Yöntem
- Ensemble Subspace Discriminant – Ensemble Yöntem
- Ensemble Subspace KNN – Ensemble Yöntem
- Ensemble RUSboosted Trees – Ensemble Yöntem

*Karar ağacı yöntemi:* Söz konusu veri setinin yapısına göre ağaç yapıları şeklinde sınıflandırma ve regresyon modelleri oluşturmaktadır.

*Destek vektör makinası (Linear):* Düzlem üzerine yerleştirilmiş noktaları ayıran doğru çizgiler çizer ve doğrunun iki sınıfının içinde noktaların maksimum uzaklıkta olmasını hedefler.

*Destek vektör makinası (Gaussian):* Sonsuz boyutlu destek vektör makinalarını bulur ve her noktanın belirli bir noktaya ne kadar benzer olduğunu hesaplamak için normal dağılım kullanır.

*En yakın k-komşu algoritması:* Komşularına bakan ve tahmin yapan bir algoritmadır. Komşuluk algoritmasında, benzer olan şeyler birbirine yakındır ve sayımı geçerlidir.

*Ensemble yöntem:* Bu öğrenme, bir modelin tek bir öğrenciyle eğitilmesi yerine birden çok öğrenciyle model oluşturmayı amaçlayan öğrenmedir. Amaç, modellerin birlikte daha doğru kararlar verebilmesidir.

Ayrıca Yapay Sinir Ağları ile de sınıflama işlemi yapılmış ve bu işlem için seçilen değişkenler MATLAB “Neural Network Pattern Recognition Toolbox” ile gerçekleştirilmiştir.

## BULGULAR

İlk olarak tüm veri için temel bileşenler analizi (principal component analysis, PCA) uygulanarak ve uygulanmadan tüm sınıflandırma yöntemleri uygulanmış ve doğru sınıflandırma yüzdeleri Tablo 1’de sunulmuştur.

**Tablo 1.** Tüm veriler için yapılan sınıflama sonuçları

Sınıflandırma Yöntemi	PCA Olmadan	PCA
	Sınıflama Doğruluğu (%)	Sınıflama Doğruluğu (%)
Fine Tree	80,80	83,70
Medium Tree	78,00	83,60
Coarse Tree	71,60	80,40
Linear SVM	<b>90,30</b>	<b>87,40</b>
Quadratic SVM	90,10	86,90
Cubic SVM	88,00	87,10
Fine Gaussian SVM	76,60	85,40
Medium Gaussian SVM	89,90	87,00
Coarse Gaussian SVM	89,90	87,40
Fine KNN	82,70	82,60
Medium KNN	84,30	85,10
Coarse KNN	85,40	86,10
Cosine KNN	84,90	85,00
Cubic KNN	84,20	84,80
Weighted KNN	85,90	86,40
Ensemble Boosted Trees	86,00	86,20
Ensemble Bagged Trees	87,10	87,00
Ensemble Subspace Discriminant	87,50	87,30
Ensemble Subspace KNN	83,70	83,50
Ensemble RUSboosted Trees	83,00	84,60

Tablo 1 incelendiğinde en başarılı sınıflama yönteminin “20 fold CV” sonuçlarına göre lineer destek vektör makinaları, en başarılı yöntemin ise “Linear SVM” olduğu görülmüştür. En iyi sınıflandırma yöntemi olarak belirlenen PCA olmaksızın uygulanan Lineer SVM için değişken seçimi işlemi

seçerek ilerleme prosedürü kullanılarak gerçekleştirilmiştir. Bu yöntemde bir değişken ile başlanarak değişken sayısı gitgide arttırılır. Toplamda ev sahibi ve deplasman takımı istatistiklerine ait 12 adet değişken bulunmaktadır. Değişkenler aşağıda Tablo 2’de listelenmiştir.

**Tablo 2.** Sınıflandırmada kullanılan değişkenlerin listesi ve açıklaması

No	Değişkenler	Açıklama
1	ace1	Ev sahibi takımların servisten aldıkları sayılar
2	Servishata1	Ev sahibi takımların servis hataları
3	hücum1	Ev sahibi takımların hücumdan kazandıkları sayılar
4	blok1	Ev sahibi takımların bloktan kazandıkları sayılar
5	manşet1	Ev sahibi takımların manşet karşılama yüzdeleri
6	ace2	Deplasman takımların servisten aldıkları sayılar
7	Servishata2	Deplasman takımların servis hataları
8	hücum2	Deplasman takımların hücumdan kazandıkları sayılar
9	blok2	Deplasman takımların bloktan kazandıkları sayılar
10	manşet2	Deplasman takımların manşet karşılama yüzdeleri
11	Cinsiyet	Erkek (cinsiyet1) – Kadın (cinsiyet2)
12	Ülke	Türkiye-İtalya

Linear SVM yöntemine göre gerçekleştirilen değişken seçimi işleminin sonucunda ace1 modele ilk olarak eklendiğinden kazancı elde edilemese de tek başına maçların %65,4 doğru sınıflandırma yüzdesi ile maçların sınıflanmasına olanak sağlamaktadır. Diğer değişkenlerin modele etki yüzdeleri Tablo 3’te verilmiştir.

**Tablo 3.** Değişkenlerin yüzdelik etkisi

No	Değişkenler	Modelde Yer Alma	Etki
1	ace1	1	
2	servishata1	1	0,20%
3	hücum1	1	5,10%
4	blok1	1	2,30%
5	manşet1	1	0,30%
6	ace2	1	4,80%
7	servishata2	0	-0,10%
8	hücum2	1	8,40%
9	blok2	1	2,20%
10	manşet2	0	-0,10%
11	Cinsiyet	0	-0,20%
12	Ülke	1	0,50%

Tablo 3’e göre servishata2 manşet2 ve Cinsiyet değişkenleri sınıflandırma performansına katkı sağlamadığı için modelden dışlanmıştır. En yüksek sınıflandırma kazancı hücum2, hücum1 ve ace2 değişkenleri ile sağlanmıştır. İkinci kısımda cinsiyete göre tüm müsabaka istatistikleri ayrılarak müsabaka sonuçlarının sınıflandırılması iki farklı şekilde uygulanmıştır. Cinsiyet göre ayrı ayrı

incelemelerde Lineer SVM en iyi yöntem olduđundan bu yöntemle göre gerekleřtirilen deđiřken seme iřleminin sonuları Tablo 4'da sunulmuřtur.

**Tablo 4.** Deđiřkenlerin yzdelik etkisi

No	Deđiřkenler	Cinsiyet=1 (erkek)		Cinsiyet=2 (kadın)	
		Modelde Yer Alma	Etki	Modelde Yer Alma	Etki
1	ace1	1		1	
2	servishata1	1	0,50%	0	-0,10%
3	hücum1	1	3,50%	1	3,60%
4	blok1	1	4,20%	1	3,00%
5	manřet1	1	0,40%	1	1,00%
6	ace2	1	5,20%	1	4,60%
7	servishata2	0	-0,10%	0	0,00%
8	hücum2	1	7,10%	1	8,60%
9	blok2	1	10,20%	1	0,90%
10	manřet2	1	0,20%	0	-0,20%
11	Ülke	1	0,30%	0	0,00%

Tablo 4'te Cinsiyete göre baktığımızda cinsiyet1 için sadece servishata2 deđiřkeni modelden dıřlanırken, cinsiyet2 için servishata1, servishata2, manřet ve Ülke deđiřkenleri modelden dıřlanmaktadır. Cinsiyet1 için ma sonucuna en ok etki eden deđiřkenin blok2 ve hücum2, Cinsiyet 2 için en yüksek etkiye sahip deđiřken hücum2 ve ace2 olduđu görölmektedir. Üüncü olarak Türkiye ve İtalya olmak üzere ölkelere göre tüm müsabaka istatistikleri ayrılarak müsabaka sonularının sınıflandırılması iki farklı řekilde uygulanmıştır. Ülkelere göre ayrı ayrı incelemelerde Lineer SVM en iyi yöntem olduđundan bu yöntemle göre gerekleřtirilen deđiřken seme iřleminin sonuları Tablo 5'da sunulmuřtur.

**Tablo 5.** Deđiřkenlerin yzdelik etkisi

No	Deđiřkenler	Türkiye		İtalya	
		Modelde Yer Alma	Etki	Modelde Yer Alma	Etki
1	ace1	1		1	
2	servishata1	1	0,10%	1	0,40%
3	hücum1	1	3,70%	0	0,00%
4	blok1	1	4,10%	1	2,90%
5	manřet1	1	0,30%	1	0,20%
6	ace2	1	4,20%	1	5,50%
7	servishata2	0	0,20%	1	0,50%
8	hücum2	1	10,00%	1	9,70%
9	blok2	1	1,60%	1	2,50%
10	manřet2	0	-0,10%	0	-0,10%
11	Cinsiyet	0	-0,40%	0	0,00%



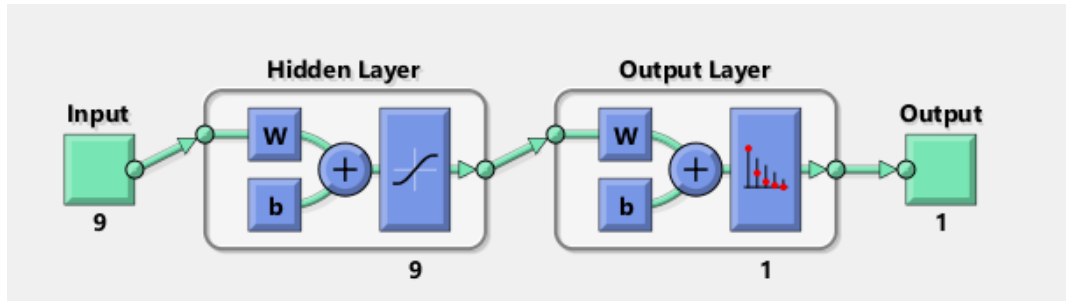
Tablo 5'e bakıldığında Türkiye için gerçekleştirilen değişken seçme işleminde servishata2, manşet2 ve Cinsiyet değişkenleri modelden dışlanmıştır. İtalya için gerçekleştirilen değişken seçme işleminde ise hücum1, manşet2 ve Cinsiyet değişkenleri modelden dışlanmıştır. Türkiye için maç sonucuna en çok etki eden değişken hücum2 ve ace2, İtalya için de en çok maç sonucuna etki eden değişkenler ise hücum2 ve ace2 olmuştur.

Yapay sinir ağları ile sınıflandırma işlemi seçerek ilerleme yöntemi ile seçilen ve aşağıda verilen değişkenler için Matlab 2018b sürümü "Neural Network Pattern Recognition Toolbox" ile gerçekleştirilmiştir ve bu versiyon k fold cross validation yönteme izin vermemektedir. Bu toolbox k kat çapraz doğrulama yerine bir kısmını dışarda bırakma (holdout) yöntemini kullanmaya izin vermektedir. Bu nedenle YSA çözümünde k kat çapraz doğrulama tercih edilmemiştir. Eğitim, Doğrulama ve Ölçek kümeleri sırasıyla %75, %15 ve %15 olarak alınmıştır. Gizli tabaka birim sayısı girdi sayısına eşit olarak 9 seçilmiştir (Tablo 6). Kullanılan YSA mimarisi Şekil 1'de verilmiştir. YSA'nın gizli tabaka birimlerinde lojistik aktivasyon fonksiyonu kullanılırken çıktı tabakasında "softmax" aktivasyon fonksiyonu kullanılmıştır. Tek gizli tabakalı mimariye sahip bu YSA'nın eğitimi Levenberg Marquardt algoritması ile Matlab'da gerçekleştirilmektedir.

**Tablo 6.** YSA için kullanılan değişkenler

ace1
servishata1
hücum1
blok1
manşet1
ace2
hücum2
blok2
Ülke

YSA'nın test sonucu dikkate alınırsa doğru sınıflama yüzdesi 89.1% olmakta ve SVM'nin bu 9 değişken ile doğru sınıflama yüzdesinden (88.9%) daha başarılı sınıflama sonucu ürettiği görülmektedir. Ancak sınıflama yüzdeleri arasındaki fark oldukça düşüktür.



**Şekil 1.** Kullanılan YSA'nın mimari yapısı

## TARTIřMA VE SONUÇ

Bu alıřmada sınıflandırma yöntemlerinin performansının deęişkenler arasındaki ilişkilerden etkilenebileceęi düşünülerek temel bileşenler analizi bir veri ön işleme teknięi olarak kullanılmıştır. İlk olarak tüm veri için temel bileşenler analizi (principal component analysis, PCA) uygulanarak ve uygulanmadan tüm sınıflandırma yöntemleri uygulanmış ve en başarılı sonuçların lineer destek vektör makinaları olduęu görülmüştür. En başarılı yöntem ise “Linear SVM (doęrusal sınıflandırma)” olmuştur. Sınıflama için belirlenen 12 deęişken üzerinden yapılan analiz sonuçlarından Linear SVM (doęrusal sınıflandırma)’de ev sahibi ve deplasman ma kazanma oranlarının doęru tahmini PCA olmadan %90,30, PCA ile %87,40 olarak belirlenmiştir. Genel analiz sonuçlarına bakıldığında Linear SVM’de deplasman takımı servis hataları ve manşet karřılama yüzdeleri ile birlikte cinsiyet deęişkenleri sınıflandırmaya katkı sağlamamıştır. En yüksek etkiyi ev sahibi ve deplasman hücumları ile deplasman takımlarının servisten direkt olarak aldıęı sayılar sağlamıştır. Bunun yanı sıra ev sahibi takımın servisten direkt aldıęı sayılar da tek başına modelde %65,4 oranına sınıflandırmaya olanak sağlamaktadır. Her iki takımın yaptıęı blok sayıları da sınıflandırmaya %2’lik bir katkı yapmıştır.

Sınıflandırma cinsiyete göre ayrı ayrı yapıldığında erkeklerde PCA analizi uygulayarak ve PCA olmadan en iyi yöntemin yine Linear SVM olduęu, kadınlarda ise PCA olmadan Linear SVM, PCA ile “Ensemble Bagged Trees (toplu torbalı ağalar)” olduęu görülmüştür. Farklı deęişkenlerin cinsiyete göre sınıflandırmaya ayrı ayrı eklenmesi sonucunda, erkeklerde sadece deplasman takımlarının servis hataları, kadınlarda ise her iki takımın servis hataları ile deplasman takımının manşet yüzdesiyle, ülke deęişkenleri modelden dıřlanmıştır. Erkek malarına bakıldığında ma soncuna en çok etki eden deęişkenin deplasman takımının hücum blok ve direkt servis sayıları olduęu görülmektedir. Bunu ev sahibi takımın hücum ve blok sayıları takip etmektedir. Kadınlara bakıldığında ma soncuna etki eden en büyük deęişken deplasman takımının hücum ve direkt servis sayıları olduęu bunu da ev sahibinin hücum ve blok sayıları takip etmektedir.

Dięer bir sınıflama yöntemi olarak ülkelere göre bakıldığında, PCA ile ve PCA olmaksızın yapılan sınıflamada hem Türkiye hem de İtalya’da en başarılı yöntemin yine Linear SVM olduęu görülmüştür. Farklı deęişkenleri ülkeye göre sınıflamaya tek tek eklediğimizde Türkiye için deplasman takımlarının servis hataları ve manşet yüzdeleri ile cinsiyet faktörü sistemden dıřlanmıştır. İtalya için ise ev sahibi takımların hücumları, deplasman takımlarının manşetleri ile cinsiyet faktörleri sistemden dıřlanmıştır. Türkiye ve İtalya için ma sonuçlarına en çok etki eden deęişkenin deplasman takımlarının hücumları olduęu görülmektedir.

Yapay sinir ağları ile yapılan sınıflandırma işlemi seçilen 9 deęişken için Matlab Neural Network Pattern Recognition Toolbox” ile gerçekleştirilmiş olup, eğitim, doęrulama ve ölçek kümeleri sırasıyla %75, %15 ve %15 olarak alınmıştır. Gizli tabaka birim sayısı girdi sayısına eşit olacak şekilde 9 seçilmiştir. YSA test sonucuna bakıldığında sınıflama doęruluęu %89,1 olarak görülmektedir. Aynı sayıda kullanılan deęişkenlerle bu sınıflama yüzdesi SVM’de %88,9 olarak çıkmaktadır. YSA sınıflandırması SVM’den yüksek olmasına rağmen aradaki fark oldukça düşüktür.

Yapılan araştırmanın amaçlarından biri olan yapay sinir ağları (YSA) ile maç sonuçları tahmini çalışmasının sonucunda sınıflama doğruluğu %89,1 olarak bulunmuştur. Literatürde bu sonucu destekleyen çalışmalar bulunmaktadır. Aka ve arkadaşları (2020) çalışmalarında futbol branşında atılan ve yenilen goller değişkenleriyle yaptığı sıralama tahmin çalışmasında YSA ile %99 oranında doğru tahmin edilebildiğini görmüştür. Ayyıldız (2018) yaptığı çalışmasında NBA liginden rastgele seçtiği müsabakalarla yaptığı sınıflandırma çalışması %90 oranında başarılı performans göstermiştir. Değer ve Süel (2022) çalışmalarında judo branşında farklı katman ve nöronlarla kurduğu bağlantılarla oluşturduğu YSA modellemesi ile maç sonuçlarını doğru tahmin edebilmiştir.

Araştırmanın bir diğer amacı olan voleybol müsabakalarının sonuçlarına etki eden değişkenlere bakıldığında erkek ve kadın maçlarında blok sayılarının müsabaka sonucuna etkileri erkek maçlarında ev sahibi takımların %4,20, deplasman takımlarının %10,20, kadın maçlarında ev sahibi takımın %3, deplasman takımlarının %1, ülke bazında bakıldığında Türkiye'de ev sahibi takımların %4,10, deplasman takımlarının %1,60, İtalya'da ev sahibi takımların %2,90, deplasman takımlarının %2,50 olarak bulunmuştur. Marcelino ve arkadaşlarının (2008) yaptıkları çalışmada müsabakalarda blok sayılarının fazlalığı maç kazanmayı olumlu yönde etkilediği görülmüştür. Palao ve arkadaşları (2009) yaptıkları çalışmada kadın ve erkek maçlarını incelemiş ve erkek maçlarında daha fazla blok yapıldığı ve maç kazanmaya etkisinin de olduğu anlaşılmıştır.

Yapılan çalışmada ulaşılan bir başka sonuca göre servisten alınan direkt sayıların müsabakaları sınıflandırmaya olan etkileri ev sahibi takımların (erkek ve kadın) aldığı sayılar ortalama %65 oranında olanak sağlamaktadır. Yine genel tabloda deplasman takımlarının direkt servis sayılarının etkisi %4,80, erkek maçlarında deplasman takımlarının servis sayıları %5,20, kadın maçlarında %4,60, ülke bazında Türkiye'de %4,20, İtalya'da %5,50 olarak bulunmuştur. Görüldüğü gibi servisten alınan direkt sayıların oranları farklı olsa da yakın değerler içermektedir. Palao ve arkadaşları (2009) yaptıkları çalışmada erkek ve kadın maçlarında farklı servis teknikleri kullanıldığını ancak bu servislerin oyuna etki yüzdelerinin benzer olduğunu tespit etmiştir.

Bu sonuçlar doğrultusunda şu önerilerde bulunulabilir:

Analizlerden yola çıkarak oynanan müsabakalarda deplasman takımlarının oyun içindeki organizasyonları müsabakaların gidişatını ve sonucunun etkileyen en önemli faktör olarak karşımıza çıkmaktadır. Fakat deplasman takımlarının bu değişkenlerinin artıp azalması ev sahibi takımların maç içerisindeki tutumlarına da bağlıdır. Erkek maçlarında en önemli yüzdelerin deplasman takımlarının blokları olduğu bir durumda, ev sahibi takımların maç esnasında atılan servise karşı alınan manşetlerin iyi olması durumunda hücum organizasyonlarının daha verimli olacağı ve rakip takım bloklarını zor durumda bırakacağı düşünüldüğünde bu oranı arttırıp azaltmak ev sahibi takımın manşet karşılama yeteneğine de bağlı olduğunu göstermektedir. Bu sebeple ev sahibi avantajını kullanacak olan takımların rakip analizi yaparken ve rakibe göre antrenman programı belirlerken rakip blokçularının durumunu ve rakip servislerini iyi analiz edip buna göre servis karşılama ve hücum antrenmanları geliştirmeleri ve çalışmalarını bu doğrultuda şekillendirmeleri müsabakalarda daha verimli olmalarının sağlayacaktır.

Kadın maçlarına bakıldığında müsabaka sonucuna etki eden en önemli deęişkenin deplasman takımlarının hücumları olduęu bir durumda, antrenörlerin dikkat etmesi ve iyi deęerlendirmesi gereken en önemli faktörün ev sahibi avantajını kullandığında takımın attıęı servislerin ve uygulayacaęı blok defans sistemi olduęu görülmektedir. Rakip takımın analizi yapılırken, kötü manşet alan oyuncuların belirlenmesi, hangi durumlarda manşet hatası yaptıklarının belirlenmesi ve servislerin bu oyuncuların bulunduęu bölgelere ve etkili bir şekilde atılmasına yönelik çalışmalar yaptırılması, aynı zamanda rakip takım hücumcularının gelen iyi yada kötü manşetlere göre topa vuruş yönleri, topu kullanma becerileri analiz edilip blok defans sisteminin bu oyunculara göre kurulacak şekilde antrenman planlaması yapılması ve çalışılması müsabakalarda ev sahibi takımları daha avantajlı konuma getirecektir.

Sonuçların deęerlendirilmesinde literatürdeki dięer çalışmalardan farklı olarak etkili deęişkenlerin seçimi için sir forward selection yöntem olarak kullanılarak deęerlendirme yapılmıştır. Bu çalışmadaki amaca uygun olarak skor üzerindeki istatistikler önem sırasına dizilmiş, ayrıca skor üzerinde etkisiz olan istatistikler belirlenmiştir. Sonuçların bir antrenör için yararlı olabileceğini düşünmekteyiz.

**Çıkar Çatışması:** Yazarlar, makalede ele alınan konu veya materyallerle ilgili olarak bir finansal veya finansal olmayan kuruluşla herhangi bir çıkar çatışması olmadığını beyan etmektedir.

**Arařtırmacıların Katkı Oranı Beyanı:** Yazar/ların arařtırmaya katkı durumları bu bölümde açıklanmalıdır. 1. Yazar % 40, 2. Yazar %30, 3. Yazar % 30 katkı sağlamıştır. Tüm yazarlar makalenin son halini okudu ve onayladı

**Etik Kurul İzni ile ilgili Bilgiler:** Giresun Üniversitesi Sosyal Bilimler Fen ve Mühendislik Bilimleri Arařtırmaları etik kurulundan 09/11/2022 tarihli ve 28/25 sayılı etik onayı alınmıştır.

## KAYNAKLAR

- Aka, H., Aktuę, Z.B., Kılıç, F. (2020). Türkiye Süper Lig Sezon Sonu Takım Sıralamasının Geliştirilen Yapay Sinir Ağları Modeli ile Tahmin Edilmesi. *Spor ve Performans Arařtırmaları Dergisi*, 11(3), 258-268.
- Akpınar, Ö. (2018). Sigorta Sektöründe Veri Madencilięi ve Kullanım Alanları. *Dumlupınar Üniversitesi Sosyal Bilimler Dergisi* 57, 103-119.
- Ayyıldız, E. (2018). Amerikan Basketbol Ligi (NBA) Maç Sonuçlarının Yapay Sinir Ağları ile Tahmini. *Gaziantep Üniversitesi Spor Bilimleri Dergisi*, 3(1) 40-53.
- Baykal, A. (2006). Veri Madencilięi Uygulama Alanları. *D.Ü. Ziya Gökalp Eğitim Fakültesi Dergisi* 7, 95-107.
- Cortsen, K., & Rascher, D. A. (2018). The application of sports technology and sports data for commercial purposes. *The use of technology in sport: Emerging challenges*, 47-84.
- Deęer, Ö., Süel, E., (2022). Gerçek Judo Müsabaka Sonuçlarının Yapay Sinir Ağları Yöntemi Yolu ile Karşılaştırılması. *Spor Eğitim Dergisi*, 6(2) 88-98.
- FIP, (t.y). <https://www.federvolley.it/>
- Hassan, A., Schrapf, N., Tilp, M. (2017). The Prediction of Action Positions in Team Handball by Non-Linear Hybrid Neural Networks. *International Journal of Performance Analysis in Sport*, 17(293-302).
- Li, Q. Y. (2018). Appliance of Apriori Algorithm on Technical-Tactics Analysis of Volleyball. In *2018 International Conference on Virtual Reality and Intelligent Systems* (164-167).

- Marcelino, R., Mesquita, I. ve Afonso, J. (2008). The Weight of Terminal Actions in Volleyball. *Contributions of the Spike, Serve and Block for the Teams Rankings in the World League*, 8(2), 1-7.
- Millington, B., Millington, R. (2015). The datafication of everything. Toward a sociology of sport and big data *Sociology of Sport Journal* 32(140-160).
- Murathan,T., Devocioğlu, S. (2018). Veri madenciliği ve spor alanındaki uygulamaları. *Hacettepe Journal of Sport Science*, 29(147-156).
- Ohgi, Y., Kaneda, K., Takakura, A. (2014). Sensor Data Mining on the Kinematical Characteristics of the Competitive Swimming. *Procedia Engineering*, 72(829-834).
- Özekes, S. (2003). Veri Madenciliği Modelleri ve Uygulama Alanları. *İstanbul Ticaret Üniversitesi Dergisi*, 2(3), 65-82.
- Palao, J.M., Manzanares, P. ve Ortega, E. (2009). Techniques Used and Efficacy of Volleyball Skills in Relation the Gender. *International Journal of Performance Analysis in Sport*, 9(2), 281-293.
- Schumaker, R. P., Solieman, O. K., & Chen, H. (2010). Sports data mining: *The field*. In *Sports Data Mining* 26(1-13).
- Sevindik, T., Kayışlı, K., Ünlükahraman, O. (2012). Web Tabanlı Eğitimde Veri Madenciliği. *Turkish Journal of Computer and Mathematics Education* 3(3), 183-193. TVF, (t.y). <https://tvf-web.dataproject.com/>
- Zhong, N. ve Zhou, L., (1999). Methodologies for Knowledge Discovery and Data Mining. *Third Pacific-Asia Conference*, 26-28.