

Veri Madenciliği ve Spor Alanındaki Uygulamaları

Data Mining and Its Applications in Sports

Araştırma Makalesi

¹Talha MURATHAN, ²Sebahattin DEVECİOĞLU

¹Ardahan Üniversitesi, Beden Eğitimi ve Spor Yüksek Okulu, Ardahan
²Fırat Üniversitesi, Spor Bilimleri Fakültesi, Elazığ

ÖZ

Hızla gelişen teknolojiyle birlikte depolanan verilerde de çok fazla artış olmuştur. Bu kadar çok veri arasından önemli ve istenilen verilerin ayırt edilebilmesi çok önemlidir. Bunun için en etkin kullanılan yöntem veri madenciliği yöntemidir. Veri madenciliği en kısa tanımıyla depolama ortamına kaydedilmiş çok fazla veri arasından istenilen, önemli ve değerli bilgilerin ortaya çıkarılmasıdır. Yeni bir konu olmakla birlikte çok fazla uygulama alanı bulunmaktadır. En fazla uygulama yapılan alanlar tıp, biyoloji ve genetikdir. Bu çalışmanın amacı veri madenciliğinin spor alanında kullanımının incelenmesidir. Ülkemizde sporda veri madenciliğiyle ilgili çok fazla bilimsel çalışma bulunmazken, yurtdışında bu alan oldukça gelişmiştir. Spor veri madenciliği teknikleri için oldukça uygun bir alandır. Sporcular, takımlar, müsabakalar ve sezonlarla ilgili çok fazla veri bulunmaktadır. Bu verilerin analiz edilmesiyle gelecekteki müsabakaların skorları tahmin edilebilmekte, oyunda yeni stratejiler belirlenebilmekte, sporcu, antrenör, tesis, sporcu araç gereçleri seçimlerinde bu verilerden yararlanılabilmekte, başarımlar değerlendirilmesi, sporcu sakatlanma riskleri, bilet satış tahminleri ve analizleri yapılabilmektedir.

Anahtar Kelimeler

Analiz, Spor, Tahmin, Teknoloji, Veri madenciliği

ABSTRACT

With the rapidly developing technology, there has been a great increase in stored data. It is very important to be able to distinguish important and desired data from so many data. The most effective method for this is the data mining method. In its shortest definition data mining is the process of extracting desired, significant and valuable information from a lot of data stored in a storage medium. Although data mining is a new issue, there are many application areas of it. Medicine, biology and genetics are the most used application areas. The purpose of this study is to examine the use of data mining in sports. While there are not many studies about data mining in sports in our country, this field has developed considerably in abroad. Sport is a very suitable field for the techniques of data mining. There are a lot of datum about the athletes, teams, competitions and seasons. By analyzing these datum, the scores of future competitions can be predicted, new strategies can be defined in a game. Furthermore; in the selection of athletes, coaches, facilities and sports equipment these datum can be used and we can make predictions about performance evaluations, injury risks of athletes, ticket sales.

Key Words

Analysis, Sports, Forecasting, Technology, Data mining

GİRİŞ

Veri Madenciliđi

Günümüzde internet kullanımının artışıyla birlikte veri miktarı da hızla artmıştır (Cao, 2012). Bilgisayar sistemlerinde saklanabilen, veri havuzlarında toplanan ve hızla artan çok sayıda veri belli bir sistematığe göre işlenmez ve analiz edilmezse insanlar için bir anlam ifade etmez. Büyük miktarlardaki veriyi işleyebilen teknikler kullanmak önemlidir. Ancak bu şekilde birçok veri arasından amaca yönelik olanlar elde edilebilir ve kullanılabilirler (Özekes, 2003). Bu tekniklerden en önemlisi veri madenciliđidir.

2025 yılında yaklaşık 1 trilyon cihazın internete bağlanacağı söylenmekte ve birçok mesleğin robotlara yaptırılacağı öngörülmektedir (Anonim, 2015). Robotlara insan düşüncesinin mantıksal yapısı aktararak işlemler bu doğrultuda yapılacaktır. Bu şekilde bağımsız karar alınması ve rasyonel karar verilmesi sağlanacaktır. Birçok cihaz internete bağlanacak, makineler içerisindeki algoritmalar, aritmetiksel ve mantıksal yöntemlerle çalışarak bizim yerimize düşünüp bizim yerimize karar vereceklerdir. Bu da yapay zekanın ortaya çıkış nedenidir. Veri madenciliđi yapay zekanın devamında geliştirilmiş bir yöntemdir. Yapay zekanın daha yavaş işlemler içeren bir yapıda olduğu gözlenmiştir (Atalay ve Çelik, 2017). Bu nedenle yapay zekanın devamı olarak bir çok alanda veri madenciliđi yönteminin kullanıldığı görülmektedir.

Veri madenciliđi, depolama ortamlarında saklanmış çok sayıda veri arasındaki ilişkiyi inceleyerek bağlantılarını bulma; istatistiksel ve matematiksel teknikler kullanılarak gereksiz verilerin elenmesiyle anlamlı yeni korelasyonları, örüntüleri ve eğilimleri keşfetme; gizli kalmış, değerli, anlamlı ve yararlı olan veriyi çekme; ve ileriye yönelik kestirimlerde bulunmayı sağlayan veri analiz tekniğidir (Cabena ve diđ., 1998; Kalikov, 2006; Özkan, 2013). Temel amacı birçok bilgi içerisinden gizli, potansiyel ve değerli bilgileri ortaya çıkarmaktır (Cao, 2012). Veri madenciliđinde problem tanımı, verilerin hazırlanması, modelin kurulması, modelin değerlendirilmesi, modelin kullanılması ve modelin izlenmesi ba-

samakları bulunmaktadır (Shearer, 2000). Veri madenciliđi çalışmalarında istatistiksel yöntemler, bellek tabanlı yöntemler, genetik algoritmalar, yapay sinir ağları ve karar ağaçları teknikleri kullanılmaktadır.

Veri madenciliđi; verinin şekli ve miktarı, donanım, bilgisayar ağları, bilimsel hesaplamalar ve ticari eğilimlerden etkilenmektedir (Akpınar, 2000). Çok fazla verinin bulunduğu ortamlarda bazen sorunlar ortaya çıkabilmektedir. Özellikle gereksiz, boş, içeriđi sürekli deđişen dinamik veriler ve eksik veriler çoksa istatistiksel analizlerde önemli problemlere neden olmaktadır (Albayrak, 2008; Savaş ve diđ., 2012).

Veri Madenciliđi Modelleri

Veri madenciliđinde tahmin edici ve tanımlayıcı modeller kullanılmaktadır. Sonuçları bilinen verilerden yola çıkılarak bir model geliştiriliyorsa ve bu modelden yararlanarak sonuçları bilinmeyen veri kümeleri için sonuç deđerleri tahmin ediliyorsa bu modellere tahmin edici modeller denmektedir (Özekes, 2003). Karar vermeye rehberlik etmede kullanılabilecek mevcut verilerdeki örüntülerin tanımlanması şeklindeki modellere ise tanımlayıcı modeller adı verilmektedir (Özekes, 2003). Veri madenciliđi modelleri gördükleri işlevlere göre; sınıflandırma, regresyon, kümeleme olmak üzere 3 grupta incelenmektedir (Joseph, 1996; Çelikten ve diđ., 2013).

Tanımlayıcı modellerden olan sınıflandırma ve regresyon modelleri arasındaki temel fark, tahmin edilen bağımlı deđişkenin kategorik veya süreklilik gösteren bir deđere sahip olmasıdır. Ancak her iki model giderek birbirine yaklaşımakta ve bunun bir sonucu olarak aynı tekniklerden yararlanılması mümkün olmaktadır.

Sınıflandırma (classification) analizi; yeni verinin incelenmesi ve daha önceden tanımlanan ve özellikleri açıkça belirlenen bir sınıfa dahil edilmesidir. Regresyon analizi ise; iki veya daha fazla deđişken arasındaki ilişkiyi belirlemek ve bu ilişkiyi kullanarak konuyla ilgili tahminler yapma yöntemidir. Tüm verilerin daima sınıflandırılma,

kategorize edilme ve derecelendirilmeye ihtiyacı olduğundan sınıflandırma veri hazırlama aracı olarak veri madenciliğinin temelini oluşturmaktadır. Verilerin sınıflandırılması iki adımda gerçekleştirilmektedir. İlk adımda veri kümelerine uygun model belirlenir. Bu model belirlenirken veritabanından rastgele seçilen eğitim verileri kullanılır. İkinci adımda ise test verileri üzerinde sınıflandırma kuralları belirlenir. Bu kurallar test verilerine dayanarak sınanır (Özkan, 2013).

Sınıflandırma ve regresyon tekniklerinde en sık kullanılan teknik karar ağaçları yöntemidir. Karar ağaçları, basit karar verme adımları uygulanarak, çok sayıda kayıt içeren bir veri kümesini çok küçük kayıt gruplarına bölmek için kullanılan bir yapıdır (Berry ve Linoff, 2004). Her başarılı bölme işlemiyle, sonuç gruplarının üyeleri bir diğeriyle çok daha benzer hale gelmektedir. Kurulumunun ucuz olması, yorumlanmalarının kolay olması, veri tabanı sistemleri ile kolayca entegre edilebilmeleri ve güvenilirliklerinin iyi olması nedenleri ile sınıflama modelleri içerisinde en yaygın kullanıma sahiptirler (Ayık ve diğ., 2007). Ayrıca Bayes sınıflandırması, K-En yakın

komşu analizi, Diskriminant analizi, Lojistik regresyon, Yapay sinir ağları, Karar destek makinaleri, Genetik algoritmalar ve Zaman serisi analizi gibi teknikler de kullanılmaktadır.

Kümeleme (cluster) analizi; verilerin benzerliklerine bağlı olarak gruplara ayrılmasıdır. Üyelerinin birbirine çok benzediği ancak özellikleri birbirinden çok farklı olan kümelerin ayırt edilmesi ve farklı kümeler oluşturulması gerçekleştirilmektedir. Bu analizde; veritabanındaki kayıtların hangi kümelere ayrılacağı veya kümelemenin hangi değişken özelliklerine göre yapılacağı konunun uzmanı olan bir kişi tarafından belirtilebileceği gibi veritabanındaki kayıtların hangi kümelere ayrılacağını geliştirilen bilgisayar programları da yapabilmektedir (Sangün, 2007).

Veri Madenciliğinin Kullanım Alanları

Veri madenciliği multidisipliner bir alana sahiptir ve yeni olmasına rağmen birkaç yıldır çok farklı disiplinlerde veri analizlerinde aktif olarak kullanılmaktadır (Chang ve Wang, 2006). Çok fazla verinin olduğu her alanda veri madenciliği

Tablo 1. Veri madenciliğinin bazı uygulamaları

Uygulama alanı	Uygulama	Referans
Tıp	Kanser erken teşhisinde	Şık, 2014
Eğitim	Lise öğrenci performanslarının değerlendirilmesi	Yurdakul, 2015
Güvenlik	Güvenlik ve sahtekarlık analizleri	Özcan, 2014
Makine	Üç fazlı asenkron motordaki sargı spirleri arasında oluşabilecek kısa devre veya yalıtım bozuklukları ve motor milinde oluşabilecek mekanik dengesizlik hatalarının belirlenmesi	Kayaalp, 2007
Sismoloji	Deprem verilerinin analizi	Duru ve Canbay, 2007
Biyoloji	DNA veri kümesinde bulunan biyolojik sıralar üzerindeki tekrarlı örüntüler ve potansiyel motiflerin çıkartılması	Baloğlu, 2006
Bankacılık	Müşteri ilişkileri yönetimi	Savaşçı ve Tatlıdil, 2006
Ticaret	Türkiye'de bir hava yolu işletmesinin parça söküm raporları	Gürbüz ve diğ., 2009
Telekomünikasyon	Sosyal ağ kullanımı	Bozkır ve diğ., 2010
Tıp	İş kazaları ve hastalıkları	Nenonen, 2013
Spor	Performans ölçümü	Soliman, 2006
Spor	Skor tahmini	Byongho ve diğ. 2008
Spor	Sporcu seçimi	Ivankovic ve diğ. 2010

uygulamaları kullanılabilir. Özellikle veri tabanı sistemleri, veri görselliđi, yapay sinir ađları, yapay öğrenme, istatistik, makine öğrenmesi, örüntü tanıma, matematiksel modelleme, spor, pazarlama, biyoloji, bankacılık, tıp, mühendislik, sigortacılık, borsa, perakendecilik, telekomünikasyon, genetik, sađlık, kriminoloji, endüstri, istihbarat, optimizasyon vb. alanlarda kullanılabilir (İnan, 2003; Albayrak, 2008; Akgöbek ve Çakır, 2009; Savaş ve diđ., 2012; Şık, 2014). Bu alanlarda olabilecek anormallik ve sahtekârlıkların ortaya çıkarılmasında da veri madenciliđinden yararlanılmaktadır (Özcan, 2014). Tablo 1'de veri madenciliđinin kullanım alanları hakkında dünyada ve Türkiye'de yapılmıř bazı örnek çalışmalar verilmiřtir. Tabloda görüldüğü gibi kanser teřhisinden deprem verilerinin analizine kadar birçođ alanda veri madenciliđi tekniklerinden yararlanılmaktadır.

Veri Madenciliđinin Spor Alanında Kullanımı

Spor veri madenciliđi için ideal bir uygulama alanıdır, çünkü oyuncular, takımlar, oyunlar ve sezonla ilgili olduđuca fazla veri bulunmaktadır

(Cao, 2012). Her takıma, her oyuncuya, her sezona, her oyuna ait veriler toplanarak, bu verilerin istatistikleri deđerlendirilebilir, örüntü analizleri yapılabilir. Sporda biriken veri miktarı arttıkça ham veriler arasından önemli verilerin ayırt edilebilmesi amacıyla pratik yöntemler bulunmaya başlanmıřtır. Sporda veri madenciliđi konusunda ülkemizde sınırlı sayıda çalışma bulunmaktadır. Yurtdıřında ise bu konuda çok fazla çalışmaya rastlanmaktadır.

Son yıllarda yapılmıř birçođ çalışma oyun verileri, sporcuların oyun içindeki durumu, spor tavsiyeleri, spor tahminleri, spor oyun videolarından oyuncuların ve oyun içerisindeki bazı alanların alıntılanması üzerine odaklanmıřtır (Min ve diđ., 2008; Choros, 2013; Lu ve diđ., 2013; Mentzelopoulos ve diđ., 2013; Messelodi ve Modena, 2013; Zeng ve Mizuno, 2013).

Veri madenciliđinin spor sektörü içerisinde dođrudan veya dolaylı olarak iliřkili olduđu alanlar Şekil 1'de verilmiřtir. Geçmiřte yapılan spor müsabakalarının veya sporcu davranıřlarının deđerlendirilmesiyle gelecekteki müsabaka sonuçları tahmin edilebilir, yeni stratejiler belirlenebilir, takımlara sporcu, araç-gereç (sporcu



Şekil 1. Veri madenciliđinin spor sektöründeki uygulama alanları

giysi ve malzemeleri) antrenör ve teknik direktör seçimlerinde bu verilerden yararlanılabilir, sporcuların sakatlanma riskleri belirlenebilir, kısa, orta veya uzun vadeli alt yapı projeleri hazırlanabilir, bilet satış tahminleri yapılabilir. Sporda veri madenciliği uygulamalarını spor branşlarında, sporda tahmin ve bahis kavramında, başarımlar değerlendirilmesi ve sporcu seçimi ile tesisleşmede, sponsorluk ve organizasyonlarda veri madenciliği uygulamaları şeklinde alt başlıklar halinde inceleyebiliriz.

Spor branşlarında veri madenciliği uygulamaları

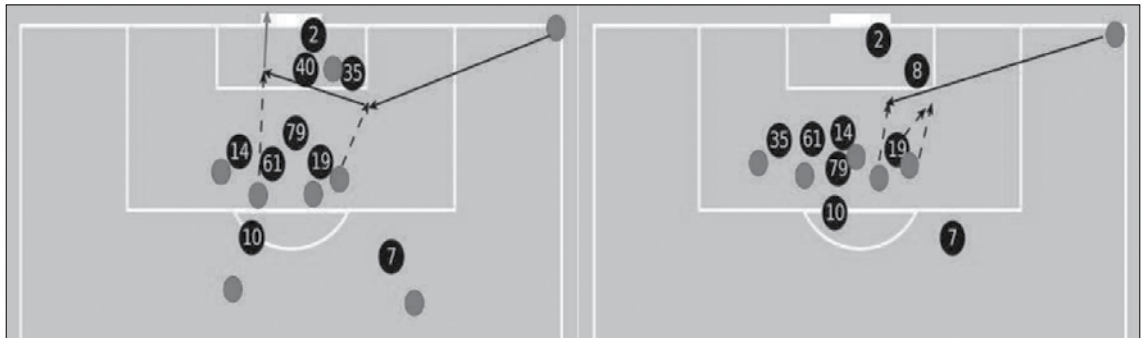
Her spor dalı için takım oluştururken sporcu seçimi kritik bir süreçtir. Bu seçimde sporcuların daha önceki karşılaşmalarda gösterdikleri performansın bilinmesi büyük önem taşımaktadır (Kalgotra ve diğ., 2013). Son yıllarda özellikle yurt dışında birçok spor takımı yetenekli yeni oyuncular keşfetmek, var olan oyuncuların eksikliklerini ortaya çıkarmak ve rakip takım durumunu analiz ettirmek amacıyla istatistikçi ve analist çalışmaktadır (Solieman, 2006). Ivankovic ve diğ. (2010) Sırbistan'da 5 sezon boyunca (2005/06, 2006/07, 2007/08, 2008/09 ve 2009/10) First B Basketbol Ligi erkek takımına ait verileri sinir ağı algoritması kullanılarak toplamışlardır. 5 sezonda oynanan toplam 890 oyuna ait verilerin analizi sonucunda önemli sonuçlara ulaşılmıştır ve bu veriler sporcu seçiminde de kullanılmıştır.

Veri madenciliği alanında birçok ülkede, futbol ve basketbol gibi daha popüler olan spor

branşlarında, yoğun veri içeren çevrimiçi ve çevrimdışı birçok program ve web sitesi bulunmaktadır. *Bu programlar yapay zeka teknolojisi ile futbol müsabakalarını video üzerinden analiz edip, saha dizilişi, oyuncu takibi, pas trafiği, rakip savunma ve hücum şekilleri, şut sayıları gibi bilgiler de ekleyerek antrenörlere sunmaktadır.*

Takımlar analitik verileri karar destek sistemleri vasıtasıyla raporlamaya dönüştürerek taktiksel hazırlıklar yapmaktadırlar. Genelde gelecekte oynanacak müsabakalar için bu değerlendirmeler kullanılmaktadır. Sporda veri madenciliği yönteminde en önemli özellik ayrıntılı görsel dokümanlardır. Şekil 2.'de görsel bir analiz örneklendirilmiştir. Türkiye Futbol Federasyonu Süper Lig İlhan Cavcav 2017-2018 Sezonunun ilk yarısının 13 ve 14. haftasında A takımının B ve C rakip takımları ile oynamış olduğu maçlarındaki sağ köşeden atılan köşe (korner) atışlarında takım dizilişi, markaj yapısı, ceza alanında bölgelerin kontrol edilmesi, yakın ve uzak direktakipleri, kale sahası içinde 2 futbolcunun kale alanı kontrolü, kontratağa yönelecek iki kişi, diğer futbolcuların da adam adama savunma şekli dizilişi görülmektedir. Bu veri yapısına bakılarak sonraki maçlarda A takımına karşı oynayacak rakip takımlar korner atışlarında taktik geliştirip avantaj sağlayabilirler. Bu veriler öngörü ve karar almada da etkili olmaktadır.

Sporda branşlarında uluslararası düzeyde kullanılmakta olan onlarca web sitesi bulunmaktadır. Örneğin futbol branşında web siteleri; fut-



Şekil 2. A takımının B ve C takımı maçlarındaki korner atışı veri karşılaştırması (Osmanlıspor. Team stats analysis, 2017)

bolcu fiziksel özellikleri, yaş, boy, kilo değerleri, transfer durumları, kazanç oranları, futbolcu ve kulüplerin piyasa değerleri, geçmişte oynanmış maç bilgi ve sonuçları, ulusal ve uluslararası üst seviye (Süper Lig, 1.Lig, Türkiye Kupası, Bundesliga, Premier League, Seri A, Ligue1, Premier Liga, Primeira Liga, La Liga vb.) liglerinin verileri, devam eden ve biten sözleşmeler, mevcut menajer ve teknik direktör verileri, ulusal ve uluslararası ligler, milli takımlar, kupalar, Avrupa, Dünya, Asya, Amerika ve Afrika kıtalarında düzenlenen turnuvalara ait verileri içerir. Türkiye’de bu tarz verilerin elde edebileceği site sayısı az sayıdadır. Türkiye Futbol Federasyonu’nun resmi web sitesinde futbolcu, teknik adam, hakem, menajer, gözlemci-temsilci, kulüpler, tüm maçlar, stad-yumlar, cezalar ve milli maç verileri bulunmaktadır.

Basketbol gibi kompleks bir oyunda, tüm maçlarda oyuncuların ribaund, asist, top çalma, blok ve turnike istatistikleri kaydedilmektedir. Bunun sonucunda çok fazla veri birikmektedir (FIBA, 2008). Amerikan Basketbol Ligi’nde takım stratejileri veri madenciliği yöntemleriyle hazırlanmaktadır.

Atletizm kategorilerinde atletlerin fiziksel durumlarını ve gelişimlerini takip ederek idman sezonunun etkinliğini hesaplamak amacıyla periyodik testler yapılmaktadır. Antrenörler daha önceden bilinen spor verilerini kullanarak sezon öncesi idmanları daha etkin hale getirebilir veya atletler üzerinde test edilen aktivitelerin etkilerini azaltabilirler. Aynı zamanda bazı organizasyonlarda spesifik şartlar altında kendi oyuncularının gösterebileceği performansı da tahmin edebilirler (Li ve Zhang, 2012).

Sporda tahmin ve bahis kavramında veri madenciliği uygulamaları

Gelecekte oynanacak oyunların skor tahminleri kolay bir iş değildir. Geleneksel olarak bu tahminler alanda uzman spor yorumcuları, televizyon ve diğer medya elemanlarında çalışan kişiler, eski sporcular, eski hakemler veya antrenörler tarafından önceki oyun sonuçlarını baz alarak, tecrübelerini, içgüdülerini ve hislerini

kullanarak yapılmaktadır. Örneğin televizyonda oyun öncesi analizlerde oyun hakkında önceden bilinenlere göre tahminler yapılmaktadır. Bu tahminler öznel iddiaları ve anekdotal kanıtları temel almaktadır. İçgüdüsel olarak yapılan öznel tahminlerin azaltılması ve daha önce oynanan oyunların değerlendirilmesiyle daha objektif olan kantitatif tahminlerin yapılması önemlidir. Özellikle profesyonel alan uzmanları, tahminlerini daha önceki sonuçları değerlendirerek yapmaktadır. Bu da özneliği azaltmaktadır (Carson ve diğ., 2014). Aynı zamanda spor tahmini yapan kişilerin bu işten büyük paralar kazanması da süreci iyice karmaşık bir hale sokmaktadır (Kalgotra ve diğ., 2013). Özellikle futbol ve basketbol gibi popüler spor branşlarında skor tahminleri günümüzde oldukça popülerdir. Bu tahminlerin yapılabilmesi için çeşitli sistemler geliştirilmiştir (Miljkovic ve diğ., 2010). Byongho ve diğ. (2008) FIFA 2002 Dünya Kupası Maçlarında veri madenciliği tekniklerinden yararlanarak çeyrek finale katılanların 8 tanesini, ikincilik alan takımı ve şampiyon olan takımı doğru tahmin edilmiştir. Yine Miljkovic ve diğ. (2010) NBA oyunlarının sonuçlarını tahmin etmek için veri madenciliği tekniklerinden yararlanmışlar ve Naive Bayes sınıflandırma yöntemini kullanmışlardır. 2009-2010 sezonu boyunca oynanan 778 oyunla ilgili veriler toplanarak değerlendirilmiştir. Maçların % 67 sinde kazanan takım hakkında teknikten yararlanılarak yapılan tahminler doğru çıkmıştır. Smith ve diğ. (2007) ise Amerikan beyzbolda kazanan takımını tahmin etmek için Naive Bayes sınıflandırma yöntemini kullanmışlardır. Sistem 1967’den 2006’ya kadar ödül kazananların % 80’ini doğru tahmin etmiştir. Yine Nunes ve Sousa (2006) Avrupa Şampiyonasının verilerinden yararlanarak görselleştirme tekniğini kullanarak veri madenciliği uygulamaları gerçekleştirmişlerdir. 2002 yılında Avrupa futbol kulübü AC Milan her oyuncunun daha önceki medikal verilerinin istatistiklerinden yararlanarak oyuncu sakatlanma tahminleri yapmak amacıyla pilot bir uygulama başlatmış ve olumlu sonuç almıştır (Flinders, 2002).

Başarım değerlendirilmesi ve sporcu seçiminde veri madenciliği uygulamaları

Veri madenciliği uygulamaları performans (başarım) değerlendirilmesinde de kullanılmaktadır. Takımların stratejik kararları başarımlarında oldukça önemlidir. Özellikle beyzbol gibi bazı branşlar veri madenciliği yöntemiyle başarımlarını belirlemede daha avantajlıdır. Atletizm yarışmalarında spor başarımlarının analizi edildiğinde başarı sırası, zaman ve skor olmak üzere üç temel yaklaşım ortaya çıkmaktadır. Bu üç ölçüm başarımların göstergesidir ve birçok spor müsabakasında atletlerin değerlendirilmesinde kullanılmaktadır (Ofoghi ve diğ., 2013). Amerikan Askeri Akademisinde okuyan askeri öğrencilerin test sonuçları arasındaki ilişkinin belirlenmesinde ve zorunlu fitness dersindeki gerçek başarımlarının değerlendirilmesinde veri madenciliği tekniklerinden yararlanıldığı bildirilmiştir (Soliman, 2006). Günay ve diğ. (2015) İstanbul'da 8-16 yaş aralığında genç kızlar basketbol takımından 166 sporcuya ait bilgileri elde etmiş, veri madenciliği teknikleriyle sınıflandırmış ve sportif becerilerine göre sporculara uygun basketbol pozisyonlarını belirlemişlerdir.

Spora başlayacak bireylerin tesadüfi olarak değil de yetenek taraması testlerinden geçirilerek, bu testlerden elde edilen verilerin anlamlandırılması ile bireylerin yatkın olduğu branşlara yönlendirilmesi veri madenciliği kullanılarak yapılabilir.

Tesisleşme sponsorluk ve organizasyonlarda veri madenciliği uygulamaları

Sanayi ve teknolojinin gelişmesi ile birlikte spor tesislerinin yapılması ve işletilmesi daha kolay bir duruma gelmiştir. Sporun yapılabilmesi ve etki alanının daha geniş tutulması için spor tesislerine ihtiyaç vardır. Günümüzde bu tesisler kamu ve özel kurum, kuruluşlar tarafından yapılmakta ve işletilmektedir. Özellikle yerel yönetimler hangi spor branşlarının halk arasında talep göreceğini belirlemek amacıyla gerek dijital gerekse yüz yüze anket ve görüşme yöntemleriyle veri toplayıp daha sonra baskın çıkan branşlarda spor tesisleri işletmeciliği veya komplekslerini hizmete

sunmaktadırlar. Yerel yönetimler bu hizmetleri vatandaşlarına sosyal belediyeçilik hizmeti olarak genelde ücretsiz bir şekilde sunmaktadırlar.

İllerdeki, bölgelerdeki ve ülkelerdeki tesislerin dağılımı, nüfusa oranlanması ve faydalanıcı sayısına göre oranlanması veri madenciliği yardımıyla yapılabilmektedir. Bununla beraber futbol sahaları, yüzme havuzları, spor salonları, tenis kortları, atletizm pistleri, halı sahalar, fitness salonlarının standartları, yapım yılları, engelli uygunlukları, toplantı odası, otopark kafeterya, sportif hizmet alanlarının çeşitliliği, iletişim bilgileri, mevcut saha bilgileri, sporcu ve seyirci kapasiteleri, kamp eğitim merkezleri, tesis tipi, işletme sahibi, tesislerin oda ve yatak sayıları, yapımı devam eden ve planlanan tesisler, spor yatırımları, spor branşlarında aktif lisanslı ve lisansız sporcu envanterleri gibi verilerin bilinmesi yapılacak tesis yatırımlarının biçimlenmesinde önem arz etmektedir.

Spor kulüplerinin sponsorluk olarak veya kendi maddi imkanları doğrultusunda branşlara göre taahhüt etikleri faaliyet alanlarının verileri ile başarı durumlarının ters veya doğru orantılı olma bilgileri ortaya çıkarılabilir.

Organizasyonlar yapılırken önceki zamanlarda yapılmış olan organizasyonların genel olarak tesis yapıları, ulaşım durumları, tesislerin kullanılabilirlik durumları, konaklama, oda ve yatak sayıları, sağlık hizmetleri, ödüllendirme işlemleri, reklam ve sponsorluk durumları, maliyetler, görsel ve teknoloji altyapısı veritabanı oluşturulması önemlidir.

Bunların dışında spor alanındaki bilimsel laboratuvar çalışmalarında elde edilen verilerin ilişkilendirilerek anlamlandırılmasında, spor hizmeti veya ürünü satın alan müşterilere yönelik olarak telefon, adres, e-posta ve tercih edilen ürünler ile alakalı anketler yapılarak elde edilen veriler nitelikli bilgilere dönüştürülerek müşterilere çeşitli ve kaliteli hizmetler sunmak amacıyla da veri madenciliğinden yararlanılabilir.

SONUÇ VE ÖNERİLER

Ülkelerin ekonomik, askeri, siyasi, sosyal vb. tüm faaliyetlerinde veri çok büyük önem arz

etmektedir. Veriler ham haliyle, işlenmemiş ve sınıflandırılmamış şekliyle çok fazla faydalı olmayabilirler. Ancak veriler bilgiye dönüştürülünce faydalı olmaktadırlar. Dünya verininin, dolayısıyla da bilginin peşinde koşmaktadır. Ülkelerin güç göstergeleri, ellerinde bulundurdukları bilgi ile ölçülmektedir.

Büyük verilerde (big data) en büyük zorluk verilerin karmaşıklığıdır. Kavram olarak büyük veriler içerisinde anlamlı ve yararlı bilgiyi seçerek ortaya çıkarmak önemlidir. Geçmişte verilerin analizi daha çok ham veriler kullanılarak analitik hesaplamalar ve istatistiksel paket programlar aracılığıyla yapılmaktaydı. Günümüzde ise verilerin tamamı herhangi bir örnekleme ihtiyaç duyulmadan ve anket kullanımına gerek kalmadan analiz edilebilmektedir.

Teknolojinin gelişmesiyle birlikte veri depolamaya sahip olan cihazlarında veri kaynakları ve kapasitelerinin artması veri madenciliğini kolaylaştırmaktadır. Küresel süreçte depolama yapan kapasite hafıza birimlerinin hacminin genişlemesiyle spor branşlarının tamamında veri akışı sağlanarak arşivleme kolay hale gelmiştir. Bu arşivlemeler yapılırken bu verilerin güvenliği de büyük önem arz etmektedir. Verileri arşivleyerek silinme ve bozulmalara karşı önlemler almak gereklidir. Verilerin dijital ortamda kayıt altında tutulurken güvenlik duvarları (firewall) ile zararlı yazılımlardan korunmaları gerekmektedir.

Veri ambarlarının kümeleme ve analitik yöntemler kullanılarak incelenmesi diğer disiplinler ile bağlantılar kurularak raporlamaya dönüştürülerek değerlendirilmesi veri madenciliği yöntemiyle mümkündür.

Tüm spor branşlarında oyuncuları ve takımları oluştururken sporcu seçimleri önemli süreçlerdir. Bu seçimler yapılırken sporculara ait kayıtlı verilerden (sezgi, algı, dikkat, teknik, serilik, kondisyon, motorik özellikler, ailelerinin genetik yapısı gibi),sporcuların sergilemiş oldukları tutum ve davranışlardan, önceki antrenman ve müsabakalarda göstermiş oldukları başarımlardan, geçmişte yaşamış oldukları sakatlıklar ve buna neden olan davranışsal risklerden fay-

dalanılarak bu verilerin bilgiye dönüştürülmesi büyük önem arz etmektedir.

Veri madenciliğinden üretilecek ve öğrenilecek bilgi ile yöneticiler, antrenörler, sporcular ve spor sektörü ile ilgili karar vericiler sektörle ilgili tüm konularda geleceğe yönelik tahmin ve öngöründe bulunarak tedbir alınmasını ve bunlara yönelik davranış geliştirilmesini sağlayacaktır. Buda istenilen hedefe ulaşma yolunda başarı sağlayacaktır.

Sporda müsabaka öncesinden toplanmış veriler ile sonuca etki eden çeşitli değişkenler vardır. Futbol branşı üzerinden değerlendirecek olursak bir takımın başarıyla ilgili pas yüzdesi, korner yüzdesi, hücum yüzdesi, faul oranı, rakip sahada veya ceza alanı içerisinde topla buluşma, pasları olumlu veya olumsuz sonuçlandırma yüzdeleri, tüm sahayı zamana yayarak kullanma oranları, toplam maç süresine kademeli olarak bakıldığında ilk yarı, ikinci yarı ve bunlar içerisindeki zaman dilimlerindeki başarımların oranları, mevkilerine göre futbolcuların başarımların göstergeleri, oyun kurmadaki isabetli isabetsiz pas yüzdeleri, sarı-kırmızı kart sayıları gibi değerler önemlidir. Bu değerlerin neticeyi ne kadar etkilediği ve hangi faktörlerin dikkate alınacağı, yapılacak hücumların taktiği, şekli, yönü gibi bilgiler ortaya çıkmaktadır. Dolayısıyla bunları önceden bilerek değerlendirmek çok önemlidir. Bu verilerden faydalanılarak rakiplere karşı üstünlük sağlanması söz konusudur.

Veri madenciliği sporda güvenliğin sağlanmasında da kullanılabilir. Büyük çapta elde edilmiş veriler içerisinde örüntüleme yapılarak istihbarati bilgi toplanması sporsal tüm alanlarda güvenlik oranını üst seviyelere çekecektir.

Veri madenciliğinin spor alanında kullanımının çeşitlendirilerek tüm spor sektörüne yayılması, hizmet sağlayan birimlere ve hizmet alanlara ileriye yönelik müsabaka, organizasyon, tahmin, risk durumları, rakip oyuncu ve takımların belirlenme özelliklerini bilerek önlem almak ve taktikler geliştirmek, sporsal her türlü seçim yapma gibi teknik ve karar alma süreçlerine büyük fayda sağlayarak başarıya katkıda bulunacaktır.

Yazışma Adresi (Corresponding Address):

Dr.Öğr.Üyesi Talha Murathan

Ardahan Üniversitesi Beden Eğitimi ve Spor
Yüksekokulu

E-posta: talhamurathan@ardahan.edu.tr

Telefon No:0505 6565324

Faks No:0.478.2117509

KAYNAKLAR

- Akgöbek Ö, Çakır F.** (2009). *Akademik Bilişim: Veri Madenciliğinde Bir Uzman Sistem Tasarımı*. Şanlıurfa: Harran Üniversitesi.
- Akpınar H.** (2000). Anonim Veri Tabanlarında Bilgi Keşfi ve Veri Madenciliği, *İstanbul Üniversitesi İşletme Fakültesi Dergisi*, 29, 1-22.
- Albayrak M.** (2008). EEG Sinyallerindeki Epilepti Form Aktivitenin Veri Madenciliği Süreci ile Tespiti. Doktora Tezi. Sakarya Üniversitesi, Fen Bilimleri Enstitüsü.
- Anonim.** (2015). <http://www.star.com.tr/ekonomi/2025-yilinda-1-trilyon-cihaz-webe-bagli-olacak-ticaret-oyoldan-donecek-haber-1055056/> 23.10.2017.
- Atalay M, Çelik E.** (2017). Büyük Veri Analizinde Yapay Zekâ ve Makine Öğrenmesi Uygulamaları. *Mehmet Akif Ersoy Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, 9(22), 155-172.
- Ayık YZ, Özdemir A, Yavuz U.** (2007). Lise Türü Ve Lise Mezuniyet Başarısının, Kazanılan Fakülte İle İlişkinin Veri Madenciliği Tekniği İle Analizi. *Atatürk Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, 10(2), 441-454.
- Baloğlu UB.** (2006). DNA Sıralarındaki Tekrarlı Örüntülerin ve Potansiyel Motiflerin Veri Madenciliği Yöntemiyle Çıkarılması, Yüksek Lisans Tezi. Fırat Üniversitesi, Fen Bilimleri Enstitüsü.
- Berry MJ, Linoff GS.** (2004). *Data Mining Techniques: For Marketing, Sales, and Customer Relationship Management* 2nd ed., Wiley, USA.
- Bigus JP.** (1996). *Data Mining With Neural Networks: Solving Business Problems From Application Development to Decision Support*. New York.
- Bozkır AS, Mazman SG, Sezer EA.** (2010). *2nd International Symposium on Information Management in a Changing World: Identification of User Patterns in Social Networks by Data Mining Techniques: Facebook Case Hacettepe University, Ankara.*
- Byungho M, Jinhuyck K, Chongyoun C, Hyeonsang E, McKay R.** (2008) A Compound Framework for Sports Results Prediction: A Football Case Study," *Knowledge-Based Systems, Elsevier Science Publishers B. V. Amsterdam, The Netherlands* 21(7), 551-562.
- Cabena P, Hadjinian P, Stadler R, Verhees J, Zanasi A.** (1998). *Discovering Data Mining: From Concept to Implementation*. New Jersey, Prentice Hall.
- Cao C.** (2012). *Sports Data Mining Technology Used in Basketball out come Prediction*. Dublin Institute of Technology for the degree of M.Sc. in Computing (Data Analytics).
- Carson KL, Kyle WJ.** (2014). *Sports Data Mining: Predicting Results for the College Football Games* *Procedia Computer Science*. 35, 710 - 719.
- Chang L, Wang H.** (2006). Analysis of Traffic Injury Severity: An Application of Non-Parametric Classification Tree Techniques *Accid. Anal. Prev.* 38,1019-1027.
- Choros K.** (2013). Temporal Aggregation of Video Shots in TV Sports News for Detection and Categorization of Player Scenes. In: *Proceedings of the ICCCI 2013*. Springer; 487-497.
- Çelikten Ö, Karacan H.** (2013). Dinamik Verilere Yönelik Karar Tahmin Mekanizması Oluşturulması. *Marmara Üniversitesi Fen Bilimleri Dergisi* 25(3), 100-114.
- Duru N, Canbay M.** (2007). *International Earthquake Symposium: Veri Madenciliği ile Deprem Verilerinin Analizi*. Kocaeli.
- FIBA** (2008). *Official Basketball Rules*. 20.12.2017, <http://www.fiba.basketball/>
- Flinders, K.** (2002). *Football Injuries are Rocket Science*. Vnunet.com. London.
- Günay SY, Arslanoğlu B, Kiriş A.** (2015). *1. International Symposium on Sport Science, Engineering and Technology: Talent Scouting in Basketball Using An Expert System Based on Fuzzy Logic and Data Mining Classification Techniques*. İstanbul.
- Gürbüz F, Özbakır L, Yapıcı H.** (2009). Türkiye'de Bir Havayolu İşletmesine Ait Parça Söküm Raporlarına İlişkin Veri Madenciliği Uygulaması. *Gazi Üniversitesi Mimarlık Mühendislik Fakültesi Dergisi*. 24(1), 73-78.
- Ivankovic Z, Rackovic M, Markoski B, Radosav D, Ivkovic M.** (2010).Appliance of Neural Networks in Basketball Scouting. *Acta Polytechnica Hungarica* 7,4.
- İnan O.**(2003). Veri Madenciliği. Yüksek Lisans Tezi. Selçuk Üniversitesi, Fen Bilimleri Enstitüsü.
- Kalgotra P, Sharda R, Chakraborty G.** (2013). Predictive Modeling in Sports Leagues: An Application in Indian Premier League. *SAS Global Forum*. 1-10.

26. **Kalikov A.** (2006). Veri Madenciliđi ve Bir E-Ticaret Uygulaması, Yüksek Lisans Tezi, Gazi Üniversitesi, Fen Bilimleri Enstitüsü.
27. **Kayaalp K.** (2007). Asenkron Motorlarda Veri Madenciliđi ile Hata Tespiti. Yüksek Lisans Tezi. Süleyman Demirel Üniversitesi, Fen Bilimleri Enstitüsü.
28. **Li Y, Zhang Y.** (2012). Application of Data Mining Techniques in Sports Training, *5th Intl. Conf. on Biomedical Engineering and Informatics (BMEI)*.
29. **Lu WL, Ting JA, Little JJ, Murphy KP.** (2013). Learning to Track and Identify Players From Broadcast Sports Videos. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 35(7), 1704-1716.
30. **Mentzelopoulos M, Psarrou A, Angelopoulou A, Rodríguez JG.** (2013). Active Foreground Region Extraction and Tracking For Sports Video Annotation. *Neural Processing Letters*. 37(1), 33-46.
31. **Messelodi S, Modena CM.** (2013). Scene Text Recognition and Tracking to Identify Athletes in Sport Videos. *Multimedia Tools and Applications*. 63(2), 521-545.
32. **Miljkovic D, Gajic L, Kovacevic A, Konjovic Z.** (2010). The use of Data Mining for Basketball Matches out Comes Prediction. *IEEE 8th International Symposium on Intelligent and Informatics*, Serbia, 309-312.
33. **Min B, Kim J, Choe C, Eon H, Ian, McKay B.** (2008). A Compound Framework for Sports Prediction: The Case Study of Football. *Knowledge-Based Systems, Elsevier Science Publishers B. V. Amsterdam, The Netherlands*. 21, 551-562.
34. **Nenonen N.** (2013). Analysing Factors Related to Slipping, Stumbling, and Falling Accidents At Work: *Application Of Data Mining Methods to Finnish Occupational Accidents and Diseases Statistics Database Applied Ergonomics*. 44, 215-224.
35. **Nunes S, Sousa M.** (2006). Applying Data Mining Techniques to Football Data From European Championships. *Actas da 1. Conferencia Metodol. Investig. Cientifica*.
36. **Ofođhi B, Zeleznikow J, MacMahon C.** (2013). Data Mining in Elite Sports. *A Review and a Framework Markus Raab Measurement in Physical Education and Exercise Science*. 17, 171-186.
37. Osmanlıspor. Team stats analysis, 13.08.2017 - 09.12.2017
38. **Özcan C.** (2014). Veri Madenciliđi ile Sahtekârlık Analizi. İstanbul Bilgi Üniversitesi Sosyal Bilimler Enstitüsü Bilişim ve Teknoloji Hukuku Yüksek Lisans Tezi, İstanbul.
39. **Özekes S.** (2003). Data Mining Models and Application Areas. İstanbul Commerce University Journal of Science. 3, 65-82.
40. **Özkan Y.** (2013). *Veri Madenciliđi Yöntemleri*, Papatya Yayınevi.
41. **Sangün L.** (2007). Temel Bileşenler Analizi, Ayırma Analizi, Kümeleme Analizleri ve Ekolojik Verilere Uygulanması Üzerine Bir Araştırma. *Çukurova Üniversitesi, Fen Bilimleri Enstitüsü, Su Ürünleri Anabilim Dalı Doktora Tezi, Adana 12s.*
42. **Savaş S, Topalođlu N, Yılmaz M.** (2012). Veri Madenciliđi ve Türkiye'deki Uygulama Örnekleri. *İstanbul Ticaret Üniversitesi Fen Bilimleri Dergisi*. 21, 1-23.
43. **Savaşçı İ, Tatlıdil R.** (2006). Bankaların Kredi Kartı Pazarında Uyguladıkları CRM (Müşteri İlişkiler Yönetimi) Stratejisinin Müşteri Sadakatine Etkisi. *Ege Akademik Bakış Dergisi*, 6, 62-73.
44. **Shearer C.** (2000). The Crisp-DM Model: The New Blueprint for Data Mining. *Journal of Data Warehousing*, 5(4), 13-23.
45. **Smith L, Lipscomb B, Simkins A.** (2007). Data Mining in Sports: Predicting Cy Young Award Winners. *Journal of Computing Sciences in Colleges*. 22 (4), 115-121.
46. **Soliman O.** (2006). Data Mining In Sports: A Research Overview. Master Project. California: University Of California Department of Managment and Informatic System.
47. **Şık MŞ.** (2014). Veri Madenciliđi ve Kanser Erken Teşhisinde Kullanımı. Yüksek Lisans Tezi. İnönü Üniversitesi Sosyal Bilimler Enstitüsü.
48. **Yurdakul S.** (2015). Veri Madenciliđi ile Lise Öğrenci Performanslarının Deđerlendirilmesi. Yüksek Lisans Tezi. Kırakkale Üniversitesi Fen Bilimleri Enstitüsü.
49. **Zeng L, Mizuno S.** (2013). On The Generalized Mirrored Scheme For Double Round Robin Tournaments in Sports Scheduling. *Asia-Pacific Journal of Operetional Research*. 30(3), 16.