



## İŞ BAŞVURULARININ MAKİNE ÖĞRENMESİ YÖNTEMLERİYLE DEĞERLENDİRİLMESİ\*

Ömer Faruk EREKEN<sup>1</sup>, Çiğdem TARHAN<sup>1</sup>

<sup>1</sup> Yönetim Bilişim Sistemleri, İktisadi ve İdari Bilimler Fakültesi, Dokuz Eylül Üniversitesi, İzmir Türkiye

### ÖZET

Bir kurum veya işletme için en önemli varlık sahip olduğu insan gücüdür. İnsan gücünün sağlam olabilmesi için en başta doğru personelin işe alınması gerekmektedir. Günümüzde kadro gereksinimleri, aday nitelikleri ve iş başvurusu sayıları artmıştır. Bu durum incelenmesi gereken veri miktarını çok yüksek boyutlara taşıyarak doğru adayı belirleme sürecini daha karmaşık ve zor hale getirmiştir. İşe alım kararı geri dönülmesi zor, uzun vadeli sonuçlar doğuran kritik bir karardır. Bu kritik kararın aynı anda birçok farklı işi takip etmek zorunda olan sınırlı sayıdaki personel tarafından verilmesi gerekmektedir. Yapay zeka yüksek boyuttaki verileri inceleyerek anlamlı çıktılar sunabilme yeteneği sayesinde yeni personel seçimine yardımcı olabilecek başlıca teknolojidir. Bu çalışma kapsamında insan kaynağı temin süreçlerinde yaşanan sorunları azaltmak ve daha kısa sürede doğru adaya ulaşılmasını sağlamak için başvuru formlarını makine öğrenmesi algoritmaları ile değerlendirerek ön eleme yapabilen bir uygulama geliştirilmiştir. İşe alım ekipleri normalde günlerce, haftalarca sürebilecek değerlendirme süreçlerini uygulama ile kısa sürelerde gerçekleştirebilecektir. Geçmişte yapılmış değerlendirmeler ile eğitilen makine öğrenmesi algoritmalarının yeni başvurular üzerinde yüksek doğruluk oranında tahminler yapabildiği gözlemlenmiştir.

**Anahtar Kelimeler:** Yapay Zeka, Makine Öğrenmesi, İnsan Kaynakları, Personel Seçimi

## EVALUATION OF JOB APPLICATIONS WITH MACHINE LEARNING METHODS

### ABSTRACT

The most significant asset for an institution or business is its manpower. For the manpower to be strong, the right staff should be hired first. Today, staff requirements, candidate qualifications and the number of job applications have increased. This situation made the process of identifying the right candidate more complex and difficult by increasing the amount of data to be analyzed to very high dimensions. The recruitment decision is a critical decision which is hard to reverse and has long-term consequences. This critical decision has to be made by a limited number of staff who have to follow many different tasks at the same time. Artificial intelligence is the primary technology that can help new staff selection, thanks to its ability to provide meaningful outputs by examining large-scale data. Within the scope of this study, in order to reduce the problems in recruitment process and to reach the right candidate in a shorter time, an application which can pre-eliminate by evaluating application forms with machine learning algorithms has been developed. With the application, recruitment teams will be able to perform the evaluation processes which can normally take days or weeks in a short

\* Bu çalışma, birinci yazarın ikinci yazar danışmanlığında hazırladığı “Yapay Zeka Tabanlı Personel Seçim Sistemi Uygulaması” başlıklı yüksek lisans tezinden üretilmiştir.

† This manuscript is produced from the “An Application Of Artificial Intelligence Based Staff Selection System” headed Master’s Dissertation prepared by the first author under the supervision of the second author.

time. It has been observed that machine learning algorithms trained with the evaluations made in the past can make predictions with high accuracy on new job applications.

**Keywords:** Artificial Intelligence, Machine Learning, Human Resources, Staff Selection

## GİRİŞ

İnsanlık tarihinin başlangıcından itibaren insanlar yeteneklerine göre iş paylaşımı yaparak belirli rolleri sahiplenmişlerdir. Bununla birlikte iş paylaşımının nasıl yapılması gerektiği, bir iş için ehil olanın kim olduğu sorunsalı ortaya çıkmıştır. Doğru kişinin seçilmesi maksadıyla tarihsel süreçte pek çok farklı yöntem izlenmiştir.

Günümüz dünyasında kurum, işletme veya şirketler için en önemli değer sahip oldukları insan gücüdür. İşleri yöneten personel gücü sağlam olmadığı takdirde sahip olunan maddi olanaklar değerlerini yitirebilmektedir. Bu nedenle iş verenler, iş yerlerine en büyük faydayı sağlayacak kişileri seçmek için çaba göstermektedirler. Her ne kadar yapılacak işlerin niteliği değişse de doğru kişinin nasıl seçileceği sorunu tüm iş kolları için varlığını korumaktadır.

Büyük ölçekli işletmelerde insan kaynakları bölümü, küçük ölçekli işletmelerde ise görevlendirilen personel veya direkt olarak yönetici tarafından gerçekleştirilen işe alım süreci, uzun ve meşakkatli bir süreçtir. Bunun yanında alınan kararlar uzun vadeli sonuçlar doğuran, geri dönüşü çok zor kararlardır.

Tüm iş kollarının profesyonelleştiği, alanında eğitim almadan, diploma, sertifika sahibi olmadan işe girmenin çok zor olduğu günümüz dünyasında işe alımdan sorumlu kişiler, başvuru yapan binlerce kişinin bilgilerini, tek tek, detaylı bir şekilde incelemek zorunda kalmaktadır. İnsan gözüyle bakıldığında tasnif etmenin çok zor olduğu yüzlerce bilgiyi okuyarak ön eleme, görüşmeye çağırma ve derecelendirme süreçlerini yürütmektedirler.

Ancak neredeyse tüm alanlarda insanoğlunun yardımına koşmaya başlayan yapay zeka ve makine öğrenmesi teknolojisi işe alım süreçlerinde de kullanıldığı takdirde haftalarca, aylarca sürebilecek olan süreçler çok kısa sürelerde gerçekleştirilebilir ve işe alım ekiplerinin enerjilerini doğru yerlere yönlendirebilir.

Bu çalışmanın amacı uzun ve zorlu bir süreç olan işe alım sürecini, yapay zeka yöntemlerini kullanarak daha hızlı ve doğru şekilde gerçekleştirmeyi sağlayacak örnek bir uygulama ortaya koymaktır. Burada ilgi odağı insan kaynağı temin sürecinin tamamen insandan bağımsız olması değil, insan kaynakları veya işe alım ekiplerinin faaliyetlerine yardımcı olacak şekilde makine öğrenmesi algoritmalarının kullanılmasıdır. Uzmanların yönlendirdiği bir makine öğrenmesi modeliyle özellikle ön eleme ve puanlandırma süreçlerinin kolaylaştırılması ve eldeki dağınık verilerden anlamlı bilgilere ulaşılması hedeflenmektedir.

## LİTERATÜR TARAMASI

İnsan kaynağı temininde karar vericilere yardımcı olabilmek için geçmişte bilgisayar tabanlı birçok çalışma yapılmıştır. Yapay zeka teknolojisinin henüz gelişmediği zamanlarda bu sistemler daha çok bilgisayarın kayıt maksatlı kullanımı, görsel analiz araçlarından faydalanma veya bilgisayarı belirli koşullara programlamayla sonuç elde etme amaçlarına yöneliktir. Günümüzde yapay zeka teknolojisinin gelişmesiyle birlikte çoğu alanda olduğu gibi insan kaynakları disiplini de entegrasyon çalışmaları hızlı bir şekilde devam etmektedir.

Aday değerlendirmeleri için kullanılan ilk sistemler daha çok adayları özelliklerine göre puanlandırma üzerine kuruluydu. Yazılımcı tarafından olası veri girişleri belirli puan ağırlıklarıyla eşleştirilmekte ve sonuçta puanlar toplanarak bir skor elde edilmekteydi. Puanlandırma sistemlerinin dezavantajı sistemin tamamen yazılımcının yönlendirmeleri doğrultusunda, belirli bir çerçevede dahilinde hareket etmesidir. Yani eğer-ise yapılarıyla kurulan temel yazılım mantığı kullanılmaktadır. Günümüz yapay zeka teknolojilerinde ise sistemin bilinç yeteneğine erişerek, bir karar verici gibi hareket etmesi hedeflenmektedir.

Yapay zekanın insan kaynağı temininde kullanımı için uzun yıllardır çalışmalar yapılmaktadır. Büyük çaplı pek çok uluslararası şirket zaman ve kaynak tasarrufu sağlayabilmek için yapay zekanın personel seçim sürecinde kullanımına yatırım yapmakta ve işe alımlarda teknolojiye faydalanmaktadır. İşe alım sürecinde yapay zekanın kullanımı belirli bir çerçevede sınırlandırılmamalıdır. Özgeçmişe göre değerlendirme, sanal oyunlar ile değerlendirme, online mülakat, sohbet robotuyla iletişim, sosyal medya değerlendirmesi gibi birçok farklı yöntem aday değerlendirmelerinde kullanılabilir.

Yapay zekanın insan kaynağı temininde kullanım alanlarından birisi adaylara ait bilgilerin yer aldığı özgeçmiş formlarının incelenmesiyle bilgi edinme yöntemidir (Strohmeier ve Piazza, 2015:164-165). Örneğin Fareportal isimli şirket iş ilanlarındaki kadro gerekliliklerini analiz ederek adayların özgeçmiş formuyla kıyaslayan “Resume Scorer” isimli bir uygulama geliştirmiştir. Böylece insan kaynakları uzmanları tek tek özgeçmiş formlarını incelemeye ayıracakları vakti en başarılı adayları incelemeye ayırmaktadır (Leong, 2018: 50-52).

Yıllık 1.8 milyon iş başvurusu alan ve bu adaylar arasından 30 binden fazla kişiyi işe alan Unilever şirketi, Pymetrics isimli işe alım uzmanlık hizmeti sağlayan şirket ile anlaşarak adayları kendi ev ortamlarında bilgisayar veya cep telefonu aracılığıyla uzaktan değerlendirebilecek bir yapay zeka sistemi kurmuştur. Öncelikle adayların bazı oyunları oynamaları beklenmektedir. İkinci aşamada adaylardan belirli soruları cevaplandırdıkları 30 dakikalık bir video mülakatı göndermeleri istenmektedir. Aşamalar makine öğrenmesi algoritmaları ile değerlendirilerek sonuçlar elde edilmektedir. Şirketin insan kaynakları bölümü yöneticisi Leena Nair yapay zeka sistemi sayesinde 70 bin personel ile saatlerce süren mülakatlar ve değerlendirme aşamalarından kurtulduklarını belirtmiştir. Örneğin şirket yapay zeka değerlendirmeleri ile milyonlarca aday arasından 3500 kişiyi değerlendirme merkezine davet etmekte ve bu adaylar arasından da 800 kişiyi işe kabul etmektedir. Sistem aynı zamanda başvuruda bulunan herkese geri dönüş yaparak şirket memnuniyetini üst seviyelerde tutmaktadır (Marr, 2018).

İnsanların kariyer durumlarını sergilediği, iş arayan ve işverenlerin buluştuğu, iş dünyasının sosyal medya platformu denilince ilk akla gelen şirket LinkedIn'dir. LinkedIn sadece sosyal medya mecrası sağlamakla kalmamış aynı zamanda kendi yapay zeka destekli işe alım sistemi “LinkedIn Recruiter”u da geliştirmiştir. LinkedIn Recruiter işverenlere aradıkları özellikler kapsamında yapay zeka teknolojilerini kullanarak sosyal medya platformu üzerinden potansiyel adayları sunmaktadır. Sonrasında işveren ve iş arayanlar arasında sosyal medya üzerinden iletişim imkanı sağlayarak iki tarafı buluşturmaktadır (LinkedIn, 2020). Sosyal medya analizini kullanan farklı çalışmalar da bulunmaktadır. Örneğin 2014 yılında makine öğrenmesi algoritmaları kullanarak yapılmış bir çalışmada sistem adayların LinkedIn profillerindeki belirli nitelikleri iş gereksinimleri ile karşılaştırarak çıkarımlar elde etmektedir. Aynı zamanda adayların blog gönderilerinde yazdıkları içerikler dil analizi ile değerlendirilmekte ve sonuçta adayın kişisel karakteristikleri belirlenmektedir (Faliaga vd., 2014: 515-528).

İnsan kaynağı temin faaliyetlerinde kullanılan yapay zeka teknolojilerinden birisi sohbet robotlarıdır. Sohbet robotları NLP (Doğal Dil İşleme) teknolojisi ile geliştirilirler ve dil kullanma yeteneğine erişmişlerdir. Sahip oldukları bu yetenek sayesinde insan kaynağı temin faaliyetlerinde mülakat veya adaylarla diğer iletişim ihtiyaçları maksadıyla kullanılabilirler. Sohbet robotları sayesinde aday başvuru formu tamamlanabilir, ayrıntılı mülakat yapılabilir, adayın şirket hakkındaki soruları cevaplanabilir veya mülakat tarihi belirlenebilir. Fakat sohbet robotlarının zaman ve kaynak tasarrufu, adaylarda stres seviyesini azaltması gibi faydalı yanlarıyla birlikte zayıf oldukları yönler de bulunmaktadır. Bunlar; empati ve mizah yönü eksikliği, eğitildiği alan (veri topluluğu) dışında bir durumla karşılaştığında yetersiz kalma ve net olmayan cevapları değerlendirememesi gibi sıkıntılardır (Joshi, 2019).

### Değerlendirme ve Problemler

Yapay zeka tabanlı işe alım sistemlerinin insan kaynağı temininde büyük faydaları bulunmaktadır. Büyük aday havuzları arasından doğru kişilerin belirlenebilmesi, yüksek orandaki verilerin incelenebilmesi, daha yüksek sayılarda adaya erişebilme, insan kaynakları uzmanlarının enerjisinin doğru alanlara yönlendirilebilmesi, uzun soluklu ve insan önyargılarının yer alabildiği mülakat süreçlerinin daha doğru değerlendirilebilmesi, güçlü sosyal medya analizlerinin mümkün olması, tüm adaylara geri besleme imkanı ve adaylar evden ayrılmadan değerlendirme yapılabilmesi gibi pek çok fayda sayılabilir. Ancak insan kaynağı temini direkt olarak insan olgusuyla ilgilendiği için hassas bir alandır. Bu sektörde gerçekleşecek olası yapay zeka hataları bir evin veya arabanın fiyat tahmininde gerçekleşen hatalardan daha ciddi sonuçlar doğurabilecektir. Günümüzde nihai kararlar her ne kadar sorumlu yöneticilere bırakılsa da sistem tarafından yapılan yanlış yönlendirmeler de hem şirket hem de bu alandaki çalışmalar için olumsuz etki oluşturabilirler.

2018 yılında basında yer alan bir habere göre büyük bir online ticaret şirketi işe alım faaliyetleri için yapay zeka teknolojisi denemeleri yapıyordu. Ancak şirket bir süre sonra yapay zekanın kadın adaylara karşı negatif bir önyargı geliştirdiğini fark etti ve teknolojinin kullanımını durdurdu. Bunun sebebi şirketin geçmişteki 10 yıllık iş başvurusu verilerini hesaba katması ve o verilerin de çoğunda erkek adayların bulunmasıydı. Makine öğrenmesi bu verilerden yanlış bir öğrenim gerçekleştirmiş ve sonuç olarak önyargılı bir yapay zeka sistemi ortaya çıkmıştı (Dastin, 2018).

Öte yandan şirketlerin yapay zeka tabanlı sistemlere geçiş yapabilmesi için gerekli alt yapı ve personel eğitimlerini tamamlamaları gerekmektedir. Hazırlık yapılmadan geçilmiş bir bilişim sistemi şirketler için yatırım ve zaman kayıplarına sebep olabilir. 2018 yılında LinkedIn tarafından yayınlanan rapora göre Kuzey Amerika'da bulunan şirketlerin sadece %22'si insan kaynakları veri analitiğini kurabilmiştir (LinkedIn, 2018: 8). Deloitte'nin 2017 Küresel İnsan Kapitali Trend Raporu'na göre şirketlerin %71'i insan analitiğini yüksek öncelikte görmekte ancak sadece %9'u kendi organizasyonları için hangi yeteneğin performans getireceğini iyi anladığını ve %8'i kullanışlı veriye sahip olduğunu ifade etmektedir (Collins vd., 2017: 97). Yapay zeka ile insan kaynağı temininde faydalı veri çok önemli bir yer tutmaktadır. Bahsi geçen raporlar ise bu alanda daha fazla yatırım ve çalışma yapılması gerektiğini kanıtlamaktadır.

### YAPAY ZEKA VE MAKİNE ÖĞRENMESİ

İnsan dünyada gördüğümüz en karmaşık ve zeki varlıktır. Ancak insanın tek başına, kendi varlığıyla ulaşabileceği noktalar sınırlıdır. Bundan dolayı tarihsel süreç içerisinde insanoğlu karşılaştığı problemlere çözüm üretmek, sınırlarını aşabilmek için belirli alanlarda kendisinden daha başarılı olabilen pek çok ürün ortaya koymuştur. Yapay zeka da insan tarafından geliştirilen, insan zekasının bir ürünü olan, makinelerin öğrenmesine ve onlara zeka kazandırılmasına dayalı bir kavramdır.

Yapay zekanın temelleri 1950 yılında Matematikçi Alan M. Turing'in yayınladığı "Hesaplayan Makineler ve Zeka" makalesine dayanmaktadır. "Makineler Düşünebilir mi?" sorusuyla makalesine giriş yapan Turing makinelerin düşünme ve zeka kabiliyetlerinin test edildiği ve kendi adıyla anılacak olan "Turing Testi"ni ortaya koymuştur (Turing, 1950: 433-460).

Makine öğrenmesi kavramı 1950'li yıllarda Arthur Samuel tarafından ortaya atılmıştır. Samuel o yıllarda hatalarından öğrenerek kendisini sürekli geliştiren dama oynayan bir bilgisayar programlamıştır (Fogg, 2017; Samuel, 1959: 211). Derin öğrenmenin tarihi ise 1943 yılında W.Pitts ve W.McCulloch tarafından insan beyninin sinir ağlarına benzer yapıda oluşturulan bilgisayar modeline dayandırılmaktadır. Ancak ilk derin öğrenme algoritmasının A.G.Ivakhnenko ve V.G.Lapa tarafından 1965 yılında ortaya konduğu söylenebilir. Yapay sinir ağları kapsamında derin öğrenme kavramının ilk defa kullanımı ise I.Aizenberg ve arkadaşları tarafından 2000 yılında olmuştur (Foote, 2017; Şeker vd., 2017: 48-49).

Makine öğrenmesi yapay zeka çatısının, derin öğrenme de makine öğrenmesi çatısının altında bulunan alt disiplinlerdir (Khanna, 2019: 457; Aydın, 2018). Yapay zeka kümesinin temsili Şekil 1'de gösterilmiştir. Bu çalışmada yapay zeka kümesi altında bulunan makine öğrenmesi yöntemleri kullanılmıştır.

### Şekil 1: Yapay Zeka Kümesi



Makine öğrenmesi bilgisayar tarafından istatistiksel yöntemler kullanılarak belli sonuçlar elde edilmesi hedeflenen bir daldır. Bilgisayarın verilerden belirli kurallar ortaya çıkarması beklenmektedir. Burada dikkat edilmesi gereken husus bilgisayara belirli kurallar programlanarak verilerden çıktılar oluşturulması makine öğrenmesi değil kurallar dahilinde bilgisayarın programlanmasıdır. Makine öğrenmesinde bilgisayar, verileri matematiksel modeller kullanarak karşılaştırır ve onlardan öğrenir. Daha sonrasında sadece veri girişi yapıldığında, çıktıları tahmin eder ya da verileri tasnif eder.

Günümüzde hayatımızda çok önemli bir yer tutan internet ve burada oluşturulan veri toplulukları ile bunlardan anlamlı sonuçlar çıkarma arayışı hız kazanmıştır. Yapay zeka artık hayatımızın her alanına girmektedir. Fiyat tahmin sistemleri, görsel algılama ve analiz, ticari tahmin ve analizler, öneri motorları ve daha pek çok alanda yapay zeka ve makine öğrenmesi kullanılmaktadır. Muhtemeldir ki önümüzdeki yıllar yapay zeka teknolojisini doğru kullanan, hayata, sektöre entegre edebilenlerin başarılı olabileceği bir dönem olacaktır. Makine öğrenmesi 3 farklı türe ayrılmaktadır. Bunlar denetimli öğrenme, denetimsiz öğrenme ve pekiştirmeli öğrenmedir (Dasgupta ve Nath, 2016: 6). Bu çalışmada denetimli öğrenme türünün sınıflandırma problemi modelleri kullanılmıştır.

### Denetimli Öğrenme

Denetimli öğrenme, girdi ve çıktıların makine öğrenmesi modeline birlikte girdiği öğrenme türüdür. Burada sistem matematiksel algoritmalar yardımıyla nitelikler (girdiler) ve çıktı (etiket) arasında bir sebep-sonuç ilişkisi kurarak bir tahmin modeli üretmekte, yani öğrenme işlemini gerçekleştirmektedir. Üretilen tahmin modeli sayesinde daha önceden görülmemiş yeni girdilerin çıktıları tahmin edilebilmektedir. Denetimli öğrenme farklı kaynaklarda farklı alt kümeler ile ifade ediliyor olsa da temel anlamda sınıfsal ve regresyon (sayısal çıktı tahmini) tipi iki probleme çözüm üretmektedir (Hastie vd., 2009: 1-10). Şekil 2'deki örnekte görüldüğü üzere burada sondaki hariç sütunlar nitelikleri ve sondaki sütun ise hedefi, çıktıyı ya da diğer bir ifade şekliyle bu verilerden ulaşılabilecek anlamlı bilgiyi temsil etmektedir.

## Şekil 2: Denetimli Öğrenmede Nitelikler ve Çıktı

|   | yas | medenidurum | egitim   | programbilgisi           | yabancidilbilgisi    | alantecrubesi | maaşbeklentisi | seyahatengeli | referans | ensonücret | vardiyalimkan | karar                      |
|---|-----|-------------|----------|--------------------------|----------------------|---------------|----------------|---------------|----------|------------|---------------|----------------------------|
| 0 | 24  | bekar       | lisans   | java,julia               | ingilizce            | 5.0           | 4000.0         | yoktur        | yoktur   | 3000.0     | çalışabilirim | Ön Görüşmeye<br>Çağrılacak |
| 1 | 32  | evli        | lisans   | java,julia               | ingilizce            | 5.0           | 5000.0         | yoktur        | vardır   | 2000.0     | çalışabilirim | Ön Görüşmeye<br>Çağrılacak |
| 2 | 27  | bekar       | önlisans | python, scala            | bilmiyor             | 3.0           | 4000.0         | yoktur        | yoktur   | 3000.0     | çalışabilirim | Elendi                     |
| 3 | 27  | evli        | önlisans | python, ofis programları | bilmiyor             | 5.0           | 4000.0         | yoktur        | yoktur   | 3000.0     | çalışabilirim | Elendi                     |
| 4 | 31  | evli        | lisans   | python, ofis programları | ingilizce, fransızca | 8.0           | 8000.0         | yoktur        | yoktur   | 2000.0     | çalışabilirim | Ön Görüşmeye<br>Çağrılacak |

Denetimli öğrenmede birinci problem tipi olan sınıflandırma insan yaşamında en sık karşılaşılan karar verme problemlerinden birisini oluşturmaktadır. Burada verilerden ulaşılmak istenen anlamlı bilgi sınıfsal bir nitelik arz etmektedir (Sathya ve Abraham, 2013: 36). Yani makine öğrenmesinden sayısal, sonsuzluk eğrisinde bir tahmin üretmesi değil kategorik bir tahmin yapması beklenmektedir. Örneğin verilere göre yapılan sahte-gerçek, geçti-kaldı, kanatlı-sürüngen gibi tahminler sınıfsal problemlere ait çıktılardır.

İkinci problem tipi devamlı, sonsuz nitelik arz eden çıktıların tahmin edilmesi problemidir. Regresyon problemlerinde sınıfsal problemlerden farklı olarak tahmin edilmek istenen değer sayısal niteliktedir. Sistemin gözlemlerden elde ettiği model ile sonsuz skalada bir değer tahmini yapması beklenmektedir. Ev kira veya satış fiyatı, araba fiyatı tahmini gibi sayısal çıktılar regresyon problemlerinin ürünüdür.

## Denetimsiz Öğrenme

Denetimsiz öğrenmede denetimli öğrenmeden farklı olarak çıktılar, yani anlamlı bilgiler yoktur. Etiket (çıktı) bilgisi olmadan makine öğrenmesi modeline giren verilerden sonuç alınmaya çalışılmaktadır. En yaygın türü olan kümeleme analizinde sistem matematiksel algoritmalar ile incelediği verileri gruplandırmaktadır (Polat, 2017: 446). Ancak sistem hangi grubun ne olduğunu bilemez sadece niteliklerden kümeler oluşturmuş olur. Sonrasında alan uzmanı kişiler kümeleri inceleyerek anlamlı bilgiye ulaşırlar.

## Pekiştirmeli Öğrenme

Pekiştirmeli öğrenme yöntemi dünyadaki canlı varlıkların öğrenme yöntemine benzemektedir. Pekiştirmeli öğrenme; en yüksek ödül bildirimini alabilmek için ne yapmak gerektiğini öğrenmektir. Eğitilen, hangi davranışları izlemesi gerektiği konusunda yönlendirilmemektedir. Onun yerine eğitilen, deneme yöntemiyle hangi davranışın en yüksek ödülü kazandıracığını keşfetmektedir. Pekiştirmeli öğrenme etiketsiz verideki saklı yapıyı bulmaya çalışan denetimsiz öğrenme ile karıştırılmamalıdır (Sutton ve Barto, 2018: 1-2).

## Sınıflandırma Problemleri Değerlendirme Ölçütleri

Makine öğrenmesinde model eğitildikten sonra belirli girdiler yapılarak tahminler alınır. Fakat model her zaman doğru tahminlerde bulunamayabilir veya regresyon problemlerinde tahmin ettiği değerler doğru değere belirli oranlarda yakınlaşır ancak tam değeri tahmin edemeyebilir. Eğitilen modellerin geliştirilmesi veya sonuçlarının istatistiksel olarak ifade edilebilmesi için modellerin analiz şekline göre değerlendirme ölçütleri bulunmaktadır. Bu çalışmada değerlendirme ölçütlerinin yalnızca bir kısmına değinilmiştir.

Sınıflandırma analizi değerlendirme ölçütlerine geçmeden önce ölçütlerde kullanılan hata matrisini incelememiz gerekmektedir. Hata matrisi sınıflandırma modellerinin tahminlerinin değerlendirilmesine olanak sağlamaktadır (Kılınç vd., 2015: 117, Sokolova ve Lapalme, 2009: 429). Hata matrisinde tahmin değerleri ile gerçek değerler kıyaslanmakta ve isimlendirilmektedir. Tablo 1'de görüldüğü üzere üst kısımda tahmin değerleri, yan tarafta ise gerçek değerler bulunmaktadır. Burada pozitif veya negatif değerler yapılan sınıflandırmanın niteliğini arz etmektedir.

**Tablo 1: Hata Matrisi**

|              |   | TAHMİN DEĞERİ                              |   |
|--------------|---|--|---|
|              |   | -  | +   |
| GERÇEK DEĞER | - | DN<br>DOĞRU NEGATİF                        | YP<br>YANLIŞ POZİTİF<br><u>I.TİP HATA</u> |
|              | + | YN<br>YANLIŞ NEGATİF<br><u>II.TİP HATA</u> | DP<br>DOĞRU POZİTİF                       |

Kaynak: Portilla, 2020.

Hata matrisini bir örnek üzerinden anlatmak daha doğru anlaşılmasını sağlayacaktır. Örneğin kişilerin seyahat vizesi alıp alamayacaklarının değerlendirildiği bir veri seti olsun. Burada gözlemdaki kişi vize alabilecek iken model doğru şekilde vize alabilir değerlendirmesinde bulunursa tabloda sağ altta yer alan doğru pozitif değerlendirme yapılmış olur. Eğer kişi vize alabilecek iken vize alamaz tahmininde bulunulursa sol altta yer alan yanlış negatif analizi yapılmıştır.

Yapılan hatalı tahminler I. ve II. Tip Hata olarak ikiye ayrılmakta ve değerlendirilen veri setinin durumuna göre bir tanesi daha kritik hata sayılabilmektedir. Kanserli hasta tahminin yapıldığı bir veri setinde kanser hastasına kanser değildir tahmini yapmak II. Tip, sağlıklı bir kişiye kanser tahmini yapmak ise I. Tip hataya girmektedir. Fakat burada II. Tip hatayı yapmış olmak kanserli bir hastayı sağlıklı olarak tahmin ettiğinden kritik bir hata sayılmaktadır ve analizcilerin bu hatadan daha çok kaçınması gerekecektir (Portilla, 2020).

### Doğruluk

Doğruluk değerlendirmesi en temel ve basit değerlendirme ölçütüdür. Doğru tahmin edilen değerlerin sayısının tüm tahminlerin sayısına bölünmesiyle doğruluk değerine ulaşılır (Nizam ve Akın, 2014: 132-133; Sokolova ve Lapalme, 2009: 430). Örneğin 100 tane gözlem verisinden 70 tane doğru sınıfsal tahmin yapıldıysa bu modelin doğruluk oranı %70'tir.

$$\text{DOĞRULUK} = \frac{\text{DOĞRU TAHMİN SAYISI (DP + DN)}}{\text{GENEL TAHMİN SAYISI}}$$

Doğruluk değerlendirmesi sınıf sayıları arasında eşit dağılım olduğu takdirde faydalıdır. Sınıfların oranı arasında büyük eşitsizlikler bulunduğu takdirde doğruluk değerlendirme ölçütünü kullanmak mantıklı değildir. Örneğin 100 tane gözlem verisinin olması gereken çıktılarının 90 tanesi kanser değildir, 10 tanesi kanser ise ve model 90 tane doğru tahmin yaptıysa bu modele başarılı demek doğruluk değerlendirmesiyle pek mümkün değildir. Çünkü model içerisine sokulan tüm değerleri kanser değildir olarak değerlendiriyor olsa ve 100 tahminin hepsine de kanser değildir çıktısı verseydi yine %90 doğruluk oranına erişmiş olacaktı (Portilla, 2020).

### Geri çağırma

Geri çağırma değerlendirme ölçütü, doğru pozitif sayısının toplam gerçek pozitif sayısına oranıdır. Toplam gerçek pozitif sayısı, doğru pozitiflerin yanlış negatiflerle toplanması ile elde edilir (Davis ve Goadrich, 2006: 235; Nizam ve Akın, 2014: 133; Sokolova ve Lapalme, 2009: 430). Pozitif değerlerin yüzde kaçının doğru tahmin edildiğini göstermektedir. Örneğin şeker hastalarının tahmin edilmeye çalışıldığı bir veri setinde geri çağırma değeri, şeker hastası olan kişilerin yüzde kaçının doğru tahmin edildiğini göstermektedir.

$$\text{GERİ ÇAĞIRMA} = \frac{\text{DOĞRU POZİTİF SAYISI (DP)}}{\text{TOPLAM GERÇEK POZİTİF SAYISI (DP + YN)}}$$

## Hassasiyet

Hassasiyet değerlendirme ölçütü, doğru tahmin edilen pozitif sayısının toplam pozitif tahmin sayısına oranıdır (Davis ve Goadrich, 2006: 233-235; Nizam ve Akın, 2014: 133; Sokolova ve Lapalme, 2009: 430). Hassasiyet, sınıflandırma modelinin gerçekten pozitif olan pozitif tahminlerinin toplam pozitif tahminine oranını vermektedir. Yani pozitif tahminlerin yüzde kaçının doğru çıktığını göstermektedir. Yine şeker hastası tahminini düşünüldüğünde burada hassasiyet şeker hastası olduğu tahmin edilen kişilerin yüzde kaçının gerçekten hasta olduğunu göstermektedir.

$$\text{HASSASİYET} = \frac{\text{DOĞRU POZİTİF SAYISI (DP)}}{\text{TOPLAM POZİTİF TAHMİN SAYISI (DP + YP)}}$$

## F1 değeri

F1 değeri, hassasiyet ve geri çağırma ölçütlerinin harmonik ortalaması alınarak elde edilen bir sınıflandırma modeli değerlendirme ölçütüdür. Bu ölçüt bir sınıfın az sayıda olduğu, oran olarak dengesiz dağılımlarda yaygın olarak kullanılmaktadır (Lipton vd., 2014: 225-227).

$$\text{HARMONİK ORTALAMA} = \frac{n (\text{ELEMEN SAYISI})}{\sum_{i=1}^n \frac{1}{x_i}}$$

$$\text{F1} = \frac{2}{\frac{1}{\text{HASSASİYET}} + \frac{1}{\text{GERİ ÇAĞIRMA}}} = 2 \times \frac{\text{HASSASİYET} \times \text{GERİ ÇAĞIRMA}}{\text{HASSASİYET} + \text{GERİ ÇAĞIRMA}}$$

## Sınıflandırma Problemlerinde Kullanılan Modeller

Sınıflandırma problemlerinin çözümünde kullanılan birçok model bulunmakla birlikte burada yalnızca çalışmada kullanılan modellere yer verilmiştir.

### Lojistik regresyon modeli

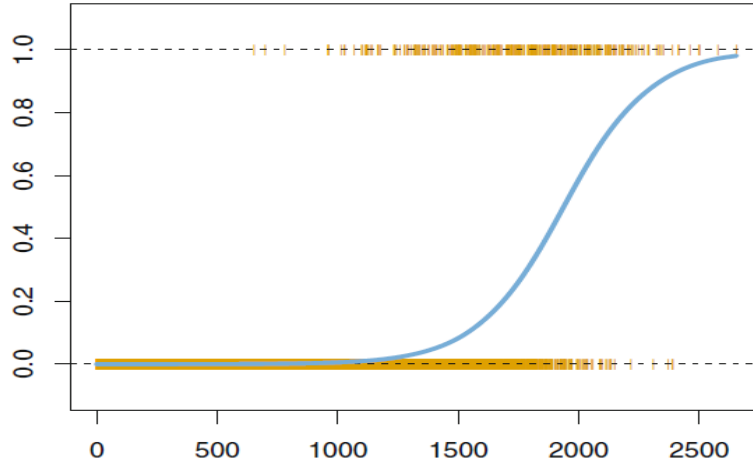
Sayısal çıktıların tahminlerinde doğrusal regresyon kullanılabilir. Ancak sınıfsal çıktılarda doğrusal regresyon kullanımının bazı sakıncaları bulunmaktadır. Bundan dolayı sınıfsal tahminler için lojistik regresyon geliştirilmiştir. Lojistik regresyonda çıktı değeri direkt olarak modellenmekten ziyade, çıktının bir sınıfa ait olma ihtimali modellenmektedir. Verilen niteliklere göre bir sınıfa ait olma ihtimali lojistik fonksiyonu kullanılarak 0-1 arasında bir değer almaktadır. Sonrasında ihtimal eşğine karar verilmektedir. Örneğin 0.5 seçildiyse, A sınıfına ait olma ihtimali 0.5 üzerinde olanlar A sınıfında, altında olanlar B sınıfındadır şeklinde sınıfsal analiz yapılmış olmaktadır. Doğrusal regresyondaki doğrunun yerini Şekil 3'te gösterilen S şekilli eğri almaktadır (James vd., 2013: 128-132).

### K en yakın komşular (KEYK) modeli

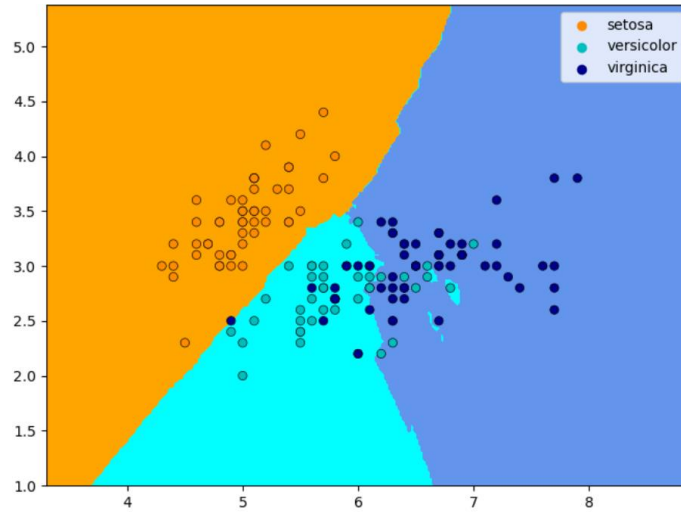
KEYK Modeli, 1966 yılında T.M.COVER ve P.E.HART tarafından bulunan "En Yakın Komşu Karar Kuralı" kullanılarak oluşturulmuştur. Bu kural sınıflandırılmamış bir örneğin daha önceden sınıflandırılmış bulunan en yakın veri kümesiyle aynı sınıfta sayılması önerisinde bulunmaktadır (Cover ve Hart, 1967: 21).

KEYK modeli makine öğrenmesinde, tahmin için girilen yeni değerleri eğitim setindeki en yakın komşularını bularak bir sınıflandırmaya tabi tutmaktadır. Birbirine yakın olan veri noktaları birbirleriyle aynı sınıfta yer alma temayülündedir. Şekil 4'te KEYK kullanılarak tasnif edilmiş 3 farklı bitki türünün makine öğrenmesi tarafından gruplandırılması ve gerçek türleri gösterilmektedir.



**Şekil 3: Lojistik Regresyon**

Kaynak: James vd., 2013: 131.

**Şekil 4: KEYK Gruplandırması**

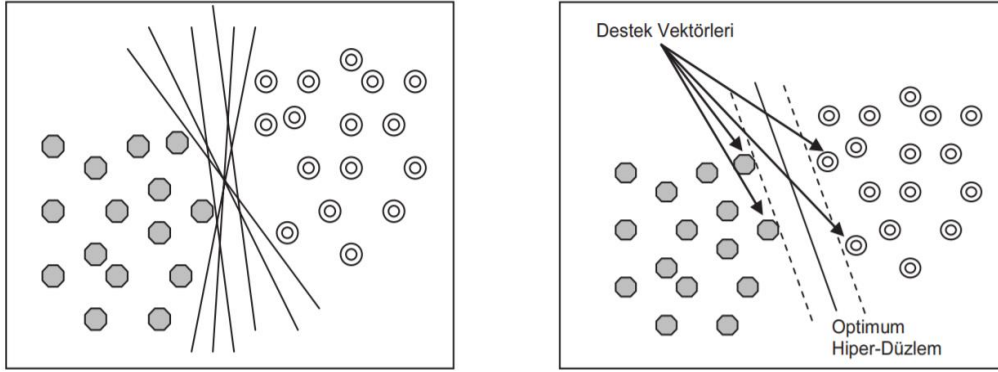
Kaynak: Scikit-learn, 2020a.

### Destek vektör makineleri (DVM) modeli

DVM'nin farklı öncüleri bulunmakla birlikte 1995 ve 1998 yıllarında Vladimir Naumovich Vapnik'in DVM'ne özel önem veren kitapları yayımlandıktan sonra metod ayrı bir ivme kazanmıştır (Steinwart ve Christmann, 2008: VII). DVM hem sınıflandırma hem de regresyon problemlerinde kullanılabilir. DVM literatürde sınıflandırma problemlerinde kullanılan modellerinin çoğuna göre daha başarılı ve etkin sonuçlar veren bir model olarak yer almaktadır. DVM sınıflandırma yapabilmek için sınıfları birbirinden ayıran optimum hiper düzlemi belirlemektedir (Ayhan ve Erdoğan, 2014: 175-178).

Daha basit ifade etmek gerekirse sınıfları birbirinden ayıran en uygun karar düzlemi bulunmaya çalışılmaktadır. Böylece yeni gelen veri noktaları düzleme olan konumlarına bakılarak sınıflandırılabilir. Şekil 5'in sol tarafında sınıfları birbirinden ayıran birçok farklı hiper düzlem çizilebileceği görülmektedir. Sağ tarafta ise sınıflara ait en yakın noktaların kendisine maksimum uzaklıkta yer aldığı optimum hiper düzlem gösterilmektedir. Optimum hiper düzlem ve veri noktaları arasındaki sınırı çizen vektörlere destek vektörleri denmektedir (Kavzaoğlu ve Çölkesen, 2010: 75-76).

### Şekil 5: DVM Hiper Düzlemi

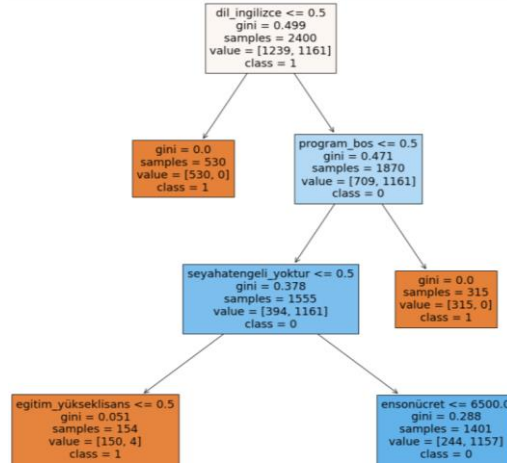


Kaynak: Kavzaoğlu ve Çölkesen, 2010:75-76

### Karar ağaçları modeli

Karar Ağaçları yüksek doğruluk oranı ve basit çalışma mantığıyla denetimli öğrenme problemlerinde yaygın olarak kullanılan başarılı bir tekniktir. Karar Ağaçları soru sorma ve cevaba göre dallanma şeklinde çalışmaktadır. Modelimizin diyagramı ters duran bir ağaç şeklindedir. Şekil 6'da görüleceği üzere en üstte kök soru aşağıya doğru da ağaç diyagramı oluşturan dallanan sorular bulunmaktadır. Amaç iki cevaplı sorularla ulaşılmak istenen çıktı sınıflarına erişmektir (Elouedi vd., 2000: 80-81).

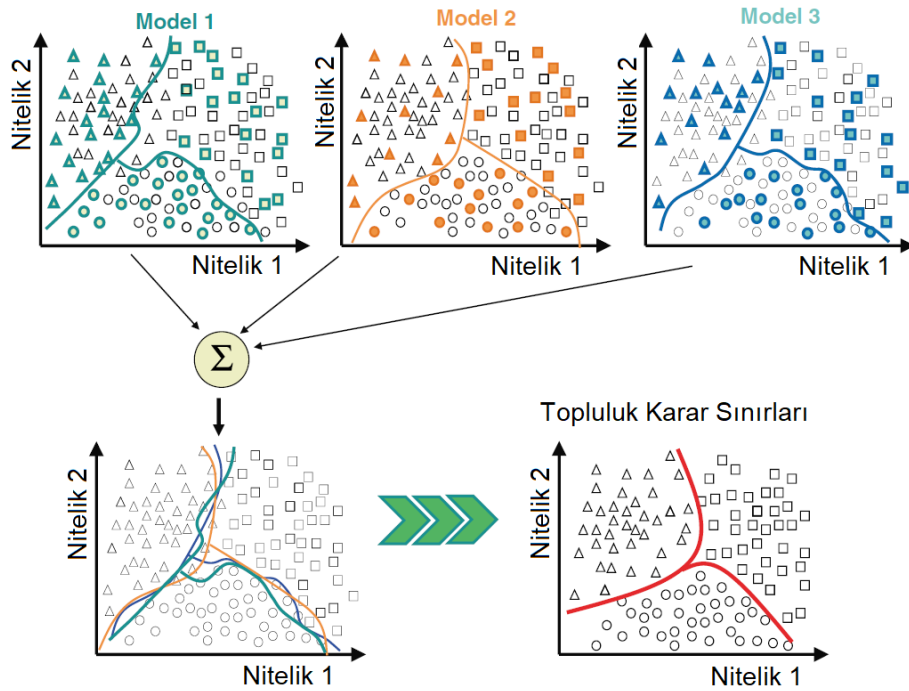
### Şekil 6: Karar Ağaçları Diyagramı



### Topluluk öğrenim metotları

Standart makine öğrenmesi modelleri bilinmeyi tanımlamak için olası fonksiyonlar (hipotez) veya karar ağaçları uzayından en iyi tahmini gerçekleştirecek fonksiyonu, ağacı bulmaya çalışmaktadır. Topluluk öğrenme algoritmaları ise veriyi anlamlandırmak için bir hipotez bulmaktansa, birçok hipotezin yer aldığı bir küme (topluluk) inşa etmekte ve sonrasında yeni bir veri için her bir hipotezin belirli ağırlıklarda tahmin yapmasıyla sonuca ulaşmaktadır (Dietterich, 2002: 405-406).

Şekil 7: Topluluk Öğrenimi Örneği



Kaynak: Polikar, 2012:3

Topluluk öğrenim metodlarında iki temel algoritmaya değinilecektir. Bunlar torbalama ve artırım algoritmalarıdır. Leo Breiman tarafından 1994 yılında geliştirilen torbalama algoritması topluluk öğrenim metodlarının en temel ve basit algoritmalarından birisi olmakla birlikte gayet etkilidir. Torbalama bir eğitim veri setinde birbirinden bağımsız sınıflandırıcıları değişimli olarak, eğitim setinden alınan örneklerle eğitir. Her bir eğitim kümesi genel veri setinden rastgele seçilmiştir. Daha sonra eğitilmiş sınıflandırıcılar basit çoğunluk oylamasıyla birleştirilir. Örneğin bir tahmin yapılacak iken tüm sınıflandırıcıların tahminleri bir oy kabul edilerek sonuca ulaşılır (Polikar, 2012:12).

Artırım, düşük tahmin başarısına sahip zayıf sınıflandırıcılar topluluğundan düşük tahmin hatasına sahip güçlü bir sınıflandırıcı elde eden bir yaklaşımdır. Torbalama gibi basit çoğunluk oylamasıyla çalışan artırım önemli bir farklılığa sahiptir. Torbalamada her eğitim kümesi genel veri setinden rastgele seçilmekteydi. Artırımında ise seçimlerde önceki sınıflandırıcılarda yanlış tahmin edilmiş veri setlerine odaklanılmaktadır (Polikar, 2012:13). Ekstrem Gradyan Artırım (XGBoost) modeli artırım topluluk öğrenimi metodunu kullanmaktadır.

### Rastgele orman modeli

Rastgele orman modeli adından da anlaşılacağı üzere aslında Karar ağaçları modelinden türetilmiştir. Karar ağaçlarında tek bir ağaç yapısı ile bir model ortaya konmaktaydı. Rastgele orman modelinde ise topluluk öğrenimi metodu olan torbalama yöntemiyle karar ağaçlarından bir orman oluşturulmakta ve tahminler ormanda bulunan ağaçların oylarıyla belirlenmektedir (Breiman, 2001: 5-6; Kılınç, Tarhan ve Aydın, 2020).

### XGBoost modeli

Makine öğrenmesinde ağaç artırımı tahmin başarısını yüksek oranlarda artırabilmektedir. Bu nedenle yaygın bir makine öğrenmesi metodudur. XGBoost da ölçeklendirilebilir uçtan uca ağaç artırımı gerçekleştiren bir algoritmadır. XGBoost, Kaggle üzerinde düzenlenen makine öğrenmesi yarışmalarında yüksek başarı ve kullanım oranıyla kendisini kanıtlamış bir sistemdir (Chen ve Guestrin, 2016: 785). Modelde her bir ağacın hatalarından öğrenerek yeni bir ağaç yapısı kurulmakta ve sonuçta en gelişmiş ağaç yapısıyla tahmin gerçekleştirilmektedir.

## VERİ VE YÖNTEM

İş başvuruları esnasında beyan edilen veriler çoğunluk itibariyle Kişisel Verilerin Korunması Kanunu kapsamında koruma altına alınan kişisel verilerdir. Bahse konu verilerin işlenebilmesi için özel şartlar haricinde kişilerin açık rızası olması zorunludur (Türkiye Büyük Millet Meclisi, 2016).

Yapay zeka sisteminin eğitilebilmesi için ise zengin içerikli iş başvurusu verilerine ihtiyaç duyulmaktadır. Gerçek verilerin temininin kanunlar ve şirket kuralları kapsamında gerçekleşmemesi nedeniyle Python programlama dili yardımıyla yazar tarafından 3000 kişilik yapay bir aday listesi oluşturulmuştur. Burada önemli olan husus sistemin verilerle eğitildikten sonra beklenen seviyede öğrenmeyi gerçekleştirip doğru çıktıları verebilmesidir. Veriler hazırlanırken gerçek hayata paralel, normalde de olması muhtemel veriler oluşturulabilmesi için üstün gayret gösterilmiştir. Temel bir örnekle açıklamak gerekirse; 1997 doğum tarihli bir kişinin doktora mezunu olması hayatın olağan akışına aykırı bir veri olacaktır. Bunun gibi, veriler doldurulurken uyumsuz, olağan dışı veri girişi olmayacak şekilde ayarlama yapılmıştır.

Veri setimiz 3000 satır ve 25 sütundan oluşmaktadır. 24 sütunluk bölüm nitelikleri, sondaki “Karar” sütunu ise çıktıyı temsil etmektedir. Veri setinde bulunan bilgiler genel anlamda Şekil 9’da girişi istenen niteliklerdir. Veri setinin oluşturulması aşamasında temel olarak Pandas kütüphanesi kullanılmıştır. Pandas, Python programlama dilinde yüksek performanslı, kullanımı kolay veri analiz araçları sağlayan açık kaynak bir kütüphanedir (McKinney vd., 2020:1; Aydın, 2019).

Anlamli bir veri seti oluşturulduktan sonra veri seti için çıktıları oluşturacak uzman karar vericiye ihtiyaç duyulmaktaydı. Bu çalışmada elimizdeki veri için karar verici yazardır. Ancak normalde bu kişi uzman personel ya da iş alımını gerçekleştirecek kişi olacaktır. İş yerlerinde insan kaynağı teminini gerçekleştirecek kişiler kararlarını aday havuzunu inceleyerek iş yeri politikası ve kadronun gerekliliklerine uygun olarak gerçekleştirmektedir.

Üzerinde çalışılan örnek işe alım süreci bir bilişim şirketinin veri bilimi uzmanı kadrosu için insan kaynağı temini olarak belirlenmiştir. Veri bilimi uzmanı sektörde aynı zamanda “Data Scientist” olarak da isimlendirilmektedir. 3000 kişiden oluşan örnek başvuru havuzunda adaylar yazar tarafından belirlenen şartlara uygun olarak değerlendirilmiştir. Şartları sağlayan kişilere mülakata çağırılacak şekilde geçti (1) etiketi verilmiş, şartları sağlamayan kişilere ise elendi (0) etiketi verilmiştir. Yapılan değerlendirme sonrasında 939 aday geçti, 2061 aday ise elendi etiketi almıştır.

İşe alım şartları her kadro ve iş kolu için değişebilmektedir. Bu şartların ne olduğunun makine öğrenmesi modeli için fazla bir önemi bulunmamaktadır. Makine öğrenmesi şartlar üzerinden değil alınan kararlar üzerinden anlamlı bilgiler edinmektedir. Örneğin; insan kaynakları uzmanı şartların yazılı olmadığı şekilde kararlar da alabilirdi. Bu durumda sistem, uzmanın verdiği kararları takip eden çıktılar verirdi.

Bu çalışmada bazı bilgiler özellikle makine öğrenmesi modelinde eğitilmemiştir. Bunlar veri setindeki cinsiyet ve medeni durum sütunlarıdır. Buradaki amaç insanlar tarafından oluşturulabilen ön yargı veya pozitif/negatif ayrımların önüne geçerek en doğru adaylara ulaşmaktır. Bunun haricinde isim, soy isim, ikamet ili, ikamet ilçesi, askerlik durumu ve doğum tarihi (yaş sütunu modele girmektedir) sütunları da karar için anlamlı bilgi içermediğinden makine öğrenmesi modeline girmeyecektir.

## Makine Öğrenmesinin Uygulanması

Elimizdeki verilerin makine öğrenmesi modeli tarafından tanınabilmesi ve üzerinde matematiksel işlemlerin gerçekleştirilebilmesi için sayısal olmayan kategorik veri sütunlarının sayısallaştırılması gerekmektedir. Bunun için Python’un Pandas kütüphanesinde bulunan “get\_dummies” metodu kullanılmıştır. Bu yöntem sayesinde kategorik veri içeren bilinen diller, mezuniyet dalı gibi bilgiler ilgili verinin bulunduğu satırın karşılığı 1, bulunmadığı satırın karşılığı 0 olacak şekilde ayarlanmaktadır (McKinney vd., 2020: 984-985).

Çalışmada makine öğrenmesi için kullanılan temel kütüphane Scikit-learn'dür. Scikit-learn orta düzey denetimli ve denetimsiz öğrenme problemleri için makine öğrenmesi algoritmaları sağlayan kullanımı kolay bir Python kütüphanesidir (Pedregosa vd., 2011: 2826). Kütüphane yüklemeleri sonrasında veri seti makine öğrenmesi için nitelik sütunları ve karar sütunu olmak üzere ikiye ayrılmıştır. Nitelikler modelin karar sütunundaki çıktıları anlamlandırmak için kullanılacağı verilerdir.

Denetimli makine öğrenmesi modellerinde mantık belirli oranlarda ikiye ayrılan etiketli verinin bir kısmının modelin eğitimi için kullanılması ve sonrasında eğitilmiş olan model tarafından daha önce görmediği veriler üzerinde tahmin gerçekleştirilmesidir. Tahmin sonrasında test verisindeki gerçek etiketler ile modelin tahminleri karşılaştırılarak modelin başarı derecesine ulaşılmaktadır.

### Model değerlendirmeleri

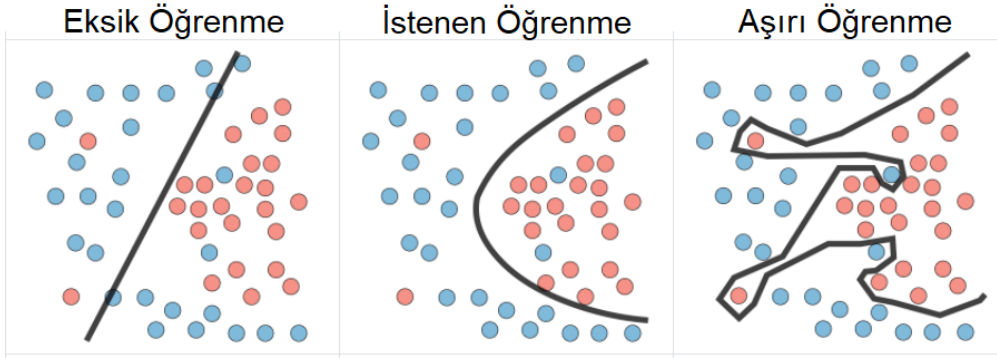
Veri setimiz üzerinde makine öğrenmesi gerçekleştirilerek elde edilen modeller ve Tablo 2'deki değerlendirme sonuçları incelendiğinde en başarılı modelin XGBoost modeli olduğu görülmektedir. XGBoost bilindiği üzere artırım topluluk öğrenim metodunu kullanmaktadır. Rastgele orman modeli de torbalama topluluk öğrenim metodunu kullanmaktadır. Ancak karar ağaçları ve rastgele orman modellerinin her ikisi de benzer çok yüksek başarı oranlarına sahip olduğundan net bir değerlendirme imkanı bulunmamaktadır.

**Tablo 2: Model Değerlendirmeleri**

|                            | Doğruluk | Hassasiyet | Geri Çağırma | F-1 Değeri |
|----------------------------|----------|------------|--------------|------------|
| <b>Lojistik Regresyon</b>  | 0.86     | 0.79       | 0.76         | 0.78       |
| <b>K En Yakın Komşular</b> | 0.82     | 0.67       | 0.85         | 0.75       |
| <b>DVM</b>                 | 0.88     | 0.82       | 0.80         | 0.81       |
| <b>Karar Ağaçları</b>      | 0.97     | 0.94       | 0.97         | 0.96       |
| <b>Rastgele Orman</b>      | 0.98     | 0.98       | 0.94         | 0.96       |
| <b>XGBoost</b>             | 1        | 0.99       | 1            | 1          |

XGBoost modelinin yaptığı tahminlerde sadece I. Tip Hata kapsamına giren normalde elenmesi gerekirken çağırılacak olarak değerlendirilmiş bir aday bulunmaktadır. Personel Seçim Sistemi Uygulamasında elenmesi gereken bir kişinin yanlışlıkla mülakata çağırılması, mülakata gelmesi gereken bir kişinin elenmesine tercih edilecek bir hatadır.

Tablo 2'deki yüksek başarı seviyeleri memnuniyet verici gözükse de model değerlendirmelerinde dikkat edilmesi gereken bir husus aşırı öğrenme gerçekleşme ihtimalidir. Aşırı öğrenme modelin elindeki veri seti üzerinden ezber mantığıyla öğrenme gerçekleştirilmesi ve bunun sonucunda görmediği tarzda bir girdide yanlış sonuçlar elde etmesidir. Şekil 40'ta görüleceği üzere aşırı öğrenme genel öğrenme eğrisinin dışına çıkarak ezberci bir patern çizmektedir. İstenen öğrenme şekli ise genel bir öğrenme paterni çizilmesidir (Amidi, 2020; Kocakafa, 2020).

**Şekil 8: Öğrenim Türleri**

Kaynak: Amidi, 2020.

Unutulmaması gereken diğer bir husus ise kullanılan verinin yapay olmasıdır. O nedenle Tablo 2'deki yüksek değerlendirme puanları tek başına, modellerin mükemmel öğrenme gerçekleştirdiğini ve sistemin halihazırındaki haliyle gerçek hayatta kullanılabileceğini göstermemektedir. Nitekim gerçek hayata ait verilerde bu kadar yüksek oranlarda başarı elde edilmesi beklenen bir durum değildir. Yazar tarafından uygulama üzerinden yapılan yeni başvuru denemelerinde özellikle XGboost modelinin çoğu tahmininde başarılı olduğu ancak nadir de olsa bazı noktalarda istenen çıktılara ulaşamadığı gözlemlenmiştir.

Sonuç olarak bu çalışmada elimizdeki veri üzerinde karar ağaçları tabanlı modellerin yüksek başarı oranı verdiği söylenebilir. Modeller ile yeni başvurular üzerinde doğru tahminler yapılabildiği görülmekte ve bu da uygulamanın insan kaynağı temin faaliyetlerinde kullanımı konusunda heyecan uyandırmaktadır. Ancak kullanılan veri yapay olduğundan daha ileri değerlendirmeler için uygulamanın gerçek hayata ait işe alım verileri üzerinde çalıştırılması ve yapılacak değerlendirmelere göre geliştirilmesi hedeflenmektedir.

**Uygulamanın Geliştirilmesi**

Yapay zeka değerlendirmeleri yapıldıktan sonra Yapay Zeka Tabanlı Personel Seçim Sistemi Uygulaması olarak adlandırılacak uygulama geliştirilmeye başlanmıştır. Uygulamada istenen şirket için web tabanlı olarak yapılacak başvuruların, aday başvurusunu tamamladığı anda kabul veya ret olarak insan kaynağı temini uzmanının önüne düşmesidir. Bunu gerçekleştirebilmek için açık kaynak bir Python kütüphanesi olan Streamlit ile web uygulaması geliştirilmesine karar verilmiştir. Streamlit özellikle makine öğrenmesi ve veri bilimi uygulama geliştiricileri tarafından kullanılan bir kütüphanedir (Streamlit, 2020).

**Başvuru girişi**

Uygulamamızda sol taraftaki seçim menüsü üzerinden seçilen Başvuru Girişi, Yönetici Girişi ve İletişim olmak üzere 3 menü bulunmaktadır. Şirkete dışarıdan başvuruda bulunacak adayların erişim yeri başvuru girişi bölümü olacaktır. Başvuru girişinde adayların girmesi gereken bilgiler yer almaktadır. Aday bilgilerini girerek başvuru tamamla butonuna bastığında veriler csv formatında kaydedilerek ileriki değerlendirmeler için saklanmaktadır. Başvuru girişi bölümünün bir kısmı Şekil 9'da görüldüğü gibidir.

**Yönetici girişi**

Yönetici girişi bölümü uygulamada makine öğrenmesi değerlendirmelerinin ve tahminlerin sergilendiği kısımdır. Yalnızca insan kaynakları uzmanları ve yöneticilerin kullanımına açıktır. Giriş belirlenen parolaya göre olmaktadır. Parola girildikten sonra makine öğrenmesi modellerimizin değerlendirilebileceği ve aynı modellere göre yeni başvuru yapan adayların tahminlerinin incelenebileceği bölümler açılmaktadır. Yanlış parola girişi yapıldığı takdirde ekranda uyarı verilmektedir.

**Şekil 9: Başvuru Girişi****Başvuru Girişi**

İsim

Soyisim

Doğum Tarihiniz

2021/03/18

İkamet İli

İkamet İlçesi

Cinsiyet

 Erkek Kız

Medenidurum

 Evli Bekar

Askerlik Durumu

 Muaf/Tamamlandı Tamamlanmadı

Eğitim Durumunuz

 Önlisans Lisans Yüksek lisans Doktora

Bildiğiniz Programlama Dillerini İşaretleyiniz

 Python R C C++ SQL Javascript Scala Julia PHP Diğer

2018-2020 Yılları Arasında Şirketimizin Düzenlediği Yapay Zeka Kursuna Katıldınız mı?

 Katıldım Katılmadım

Bildiğiniz Yabancı Dilleri İşaretleyiniz

 İngilizce Almanca Fransızca Arapça Rusça Yabancı Dil Bilmiyorum Diğer

IT Alanındaki İş Tecrübeniz Kaç Yıldır?

Maaş Beklentiniz Aylık Kaç TL'dir?

En Son Çalıştığınız Şirketteki Maaşınız Aylık Kaç TL'di?

Şekil 10: Yönetici Yan Sekme Ekranı

×

MENÜ

Yönetici Girişi ▾

Parola

👁

Giriş

İşlemi Seçiniz

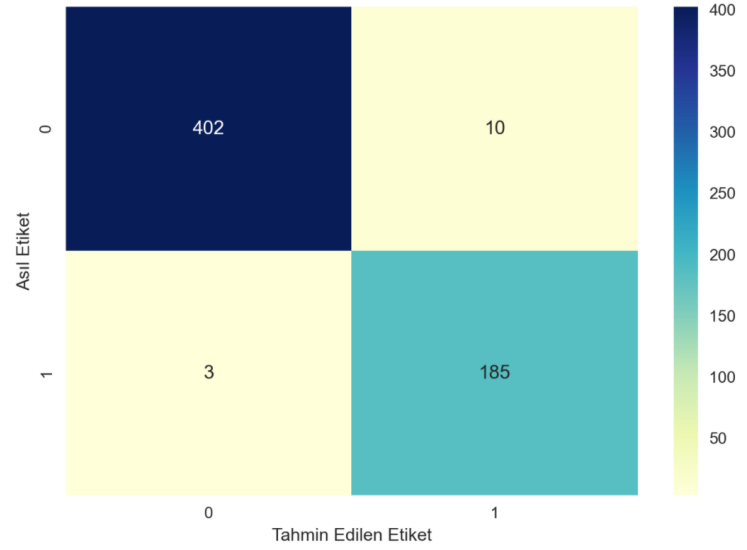
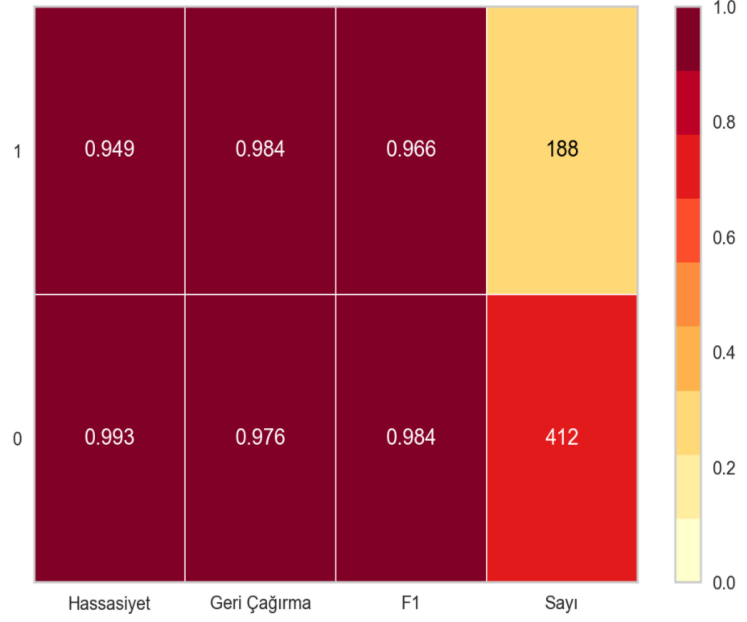
Model Puanlamalarını Görme ▾

Modeli Seçiniz

Karar Ağacı ▾

Doğruluk Oranı = 0.9783333333333334

## Sınıflandırma Raporu





Yönetici girişi bölümünde iki adet yan sekme açılmaktadır. Üstteki sekmeden model puanlarını görme veya başvuruları değerlendirme seçenekleri belirlenmektedir. Alttaki sekmede iki işlem için de çalışan model seçimi sekmesi bulunmaktadır. Model puanlarını görme sekmesinde Şekil 10'da görüldüğü gibi hassasiyet, geri çağırma ve F1 değerlendirmeleri sınıflandırma raporunda, tahminlerin durumu ise hata matrisinde görsel olarak gösterilmektedir.

Başvuru değerlendirme bölümü seçildiğinde tüm veri setiyle eğitilen modeller yeni başvurular verisi üzerinde tahminlerini yapacaktır. Model ve parametre seçimlerine göre interaktif olarak tahminler de değişebilmektedir. Model değerlendirme bölümünde en başarılı model görüldükten sonra tahmin için o model seçilebilir. Şekil 11'de karar kısmında görülen 1 değeri seçili bulunan XGBoost modelinin adayı geçti olarak, 0 ise elendi olarak değerlendirdiğini göstermektedir.

Şekil 11: Başvuru Değerlendirme Ekranı

| karar | isim | soyisim | dogumtarihi | ikametili | ikametilcesi | cinsiyet |
|-------|------|---------|-------------|-----------|--------------|----------|
| 0     | 1    | Mehmet  | 1988-05-25  | İstanbul  | Kadıköy      |          |
| 1     | 1    | Metin   | 1989-07-20  | İstanbul  | Kartal       |          |
| 2     | 1    | Meltem  | 1991-01-16  | İstanbul  | Beylikdüzü   |          |
| 3     | 0    | Ayşe    | 1995-04-26  | İstanbul  | Beylikdüzü   |          |
| 4     | 0    | Erden   | 1981-05-14  | İstanbul  | Beykoz       |          |

## SONUÇ

Kurumların sahip oldukları en kıymetli değer olan insan gücünü doğru şekilde belirleyebilmeleri için birçok çalışma yapılmaktadır. Özellikle son yıllarda bu alanda yapay zeka teknolojisinin kullanımı için çalışmalar hız kazanmıştır. Uluslararası boyutta büyük şirketler milyonlarca kişi arasından en doğru kişileri yapay zeka teknolojilerinden faydalanarak seçmektedir. Ancak teknoloji yeni kullanılmaya başlandığından hala nihai kararlarda şirket karar vericileri kontrol sahibidir. Yapay zeka teknolojisinin alanda kullanımının farklı versiyonları bulunmaktadır. Yapay zeka vasıtasıyla özgeçmiş formlarının değerlendirilmesi, sanal oyunlar ile değerlendirme, videolu online mülakat, sohbet robotlarıyla mülakat veya sosyal medya analizi kullanılan bazı yöntemlerdir.

Bu çalışma kapsamında ise yapay zeka vasıtasıyla özgeçmiş formlarının değerlendirilmesi yöntemine benzer olarak form tabanlı aday değerlendirmesi yöntemi izlenmiştir. Özgeçmiş formları değerlendirmesinde adayların veri formlarından bilgi çekerek analiz etme yöntemi izlenmektedir. Ancak özgeçmiş formları genellikle adaylar tarafından farklı formatlarda, faydalı olmayan bilgiler de ihtiva edebilecek şekilde hazırlanabilmektedir. Bundan dolayı başvuru formu tabanlı değerlendirmenin ilk değerlendirmeler için daha başarılı olacağı düşünülmektedir. İlk başta adaylardan direkt olarak hedefe yönelik bilgi talebi yapılarak sürecin hızlandırılması ve daha doğru sonuçların elde edilmesi hedeflenmiştir.

Yapay Zeka Tabanlı Personel Seçim Sistemi Uygulaması makine öğrenmesi yöntemleri kullanılarak web ara yüzü ile gerçekleştirilmiştir. Çalışmanın ilk aşamasında makine öğrenmesinde eğitilecek veri seti hazırlanmış ve adaylar karar verici yazar tarafından geçti/kaldı şeklinde iki farklı sonuçla etiketlenmiştir.

Etiketlenmiş veri seti ile Lojistik Regresyon, K En Yakın Komşular, Destek Vektör Makineleri, Karar Ağaçları, Rastgele Orman ve XGBoost denetimli makine öğrenmesi modelleri

eğitilmiştir. Değerlendirme aşamasında veri seti için karar ağaçları tabanlı modellerin daha başarılı olduğu ve topluluk öğrenim metodunu kullanan XGBoost modelinin en iyi sonucu veren sınıflandırma algoritması olduğu tespit edilmiştir. Ancak kullanılan veri yapay olduğundan daha ileri seviye analizler için gerçek hayat verileri ile çalışma yapılması gerektiği değerlendirilmiştir.

Makine öğrenmesi modelleri geliştirildikten sonra başvuru formlarının doldurulup, verilerin kaydedileceği ve modellere göre başvuru değerlendirmelerinin yapılabileceği web ara yüzünde uygulama geliştirilmiştir. Uygulamanın başvuru bölümünde aday bilgilerinin girişi yapılarak tamamlama butonuna basıldığında veriler kaydedilmektedir. Sonrasında yönetici girişi bölümünden istenen makine öğrenmesi modeline göre aday için karar tahminleri izlenebilmektedir.

Bu çalışmada insan kaynağı temininde ön eleme aşamasında değerlendirmelerde bulunulmuş ancak sistem istendiği takdirde ufak düzenlemelerle sayısal çıktılarının bulunduğu anlamlı bir veri setinde de çalıştırılabilecektir. Sayısal veriler kullanıldığı takdirde makine öğrenmesi regresyon modelleri kullanılacaktır.

Veri temizleme safhası hesaba katılmadığı takdirde geçmişte binlerce kişi üzerinde yapılmış değerlendirmeler, elimizdeki uygulama ile dakikalar içerisinde incelenerek, yeni başvuruların değerlendirilmesinde kullanılabilir.

Yönetim bilişim sistemleri, anlamsız veri topluluklarından anlamlı bilgiye ulaşmayı ve bu anlamlı bilgiyle karar vericileri doğru şekilde yönlendirmeyi hedefleyen bir disiplindir. Yapay Zeka Tabanlı Personel Seçim Sistemi Uygulaması da yöneticilerin kurumlarına yapacakları personel seçiminde hızlı ve doğru kararlar almalarına yardımcı olan bir yönetim bilişim sistemidir.

Bu çalışmanın geçmiş yıllara ait personel alımı verilerini saklayan bir şirkette, gerçek veriler üzerinde denenmesi ve alınan sonuçlara göre gerekli geliştirmelerin yapılması hedeflenmektedir. İlerleyen çalışmalarda örneklem sayısı çeşitlendirilerek modellerin öğrenme durumları konusunda daha detaylı bilgi elde edilecektir. Mülakata çağırılan adayların gelmemesi durumunda yerlerine çağırılacak yedek adayların seçimi için bir otomasyon sistemi kurulacaktır. Ön eleme safhası sonrasında, yapay zeka puanlandırması ile aday başarı sıralaması yapılması ve yapay sinir ağları kullanarak makine öğrenmesi modellerinin yanında derin öğrenme algoritmaları ile de tahminler yürütülmesi planlanmaktadır.

Sonuç olarak bu çalışma işe alım faaliyetlerinde yapay zeka ve makine öğrenmesi tabanlı sistemlerin geliştirilmesinin maddi kaynaklar ve zaman açısından büyük tasarruflar sağlayabileceğini göstermektedir. Yalnızca insan kaynağı temin süreçlerinde değil karar verme faaliyetinin gerçekleştirildiği her alanda yapay zeka teknolojilerinden faydalanılması ve bu alanlarda daha fazla çalışma yapılması gelecek adına yapılabilecek en mantıklı yatırımlardan birisi olacaktır.

## KAYNAKÇA

Amidi, A. ve Amidi, S. (2020). Machine Learning Tips and Tricks Cheatsheet. <https://stanford.edu/~shervine/teaching/cs-229/cheatsheet-machine-learning-tips-and-tricks#>, (28.12.2020).

Aydın, C. (2018). Makine Öğrenmesi Algoritmaları Kullanılarak İtfaiye İstasyonu İhtiyacının Sınıflandırılması . Avrupa Bilim ve Teknoloji Dergisi , (14) , 169-175 . DOI: 10.31590/ejosat.458613.

Aydın, C. (2019). Classification of the Fire Station Requirement with Using Machine Learning Algorithms. I.J. Information Technology and Computer Science. 1, 24-30. DOI: 10.5815/ijitcs.2019.01.03.

Ayhan, S. ve Erdoğan, Ş. (2014) Destek Vektör Makineleriyle Sınıflandırma Problemlerinin Çözümü İçin Çekirdek Fonksiyonu Seçimi. Eskişehir Osmangazi Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi. 9(1): 175-201.

Breiman, L. (2001) Random Forests. Machine Learning. 45: 5-32.

Chen, T. ve Guestrin, G. (2016). XGBoost: A Scalable Tree Boosting System. KDD'16: Proceedings Of The 22nd ACM SIGKDD International Conference On Knowledge Discovery and Data Mining (ss.785-794), Düzenleyen Association For Computing Machinery. San Francisco. 13-17 Ağustos 2016.

Collins, L., Fineman, D.R. ve Tsuchida, A. (2017). People Analytics: Recalculating The Route. Rewriting The Rules For The Digital Age 2017 Deloitte Global Human Capital Trends. (ss.97-105). Deloitte University Press.

Cover, T.M ve Hart, P.E. (1967). Nearest Neighbor Pattern Classification. IEEE Transactions On Information Theory. 13(1): 21-27.

Dasgupta, A. ve Nath A. (2016). Classification of Machine Learning Algorithms. International Journal of Innovative Research in Advanced Engineering. 3(3): 6-11.

Dastin, J. (11 Ekim 2018). Amazon Scraps Secret AI Recruiting Tool That Showed Bias Against Women. <https://www.reuters.com/article/us-amazon-com-jobs-automation-insight-idUSKCN1MK08G>, (20.12.2020)

Davis, J. ve Goadrich M. (2006) The Relationship Between Precision-Recall and ROC Curves. Proceedings Of The 23rd International Conference On Machine Learning (ss. 233-240), Düzenleyen The International Machine Learning Society. Pittsburgh. 25-29 Temmuz 2006.

Dietterich, T.G. (2002). Ensemble Learning. The Handbook Of Brain Theory and Neural Networks. (ss. 405-408). Cambridge: The MIT Press.

Elouedi, Z., Mellouli K. ve Smets P. (2000) Classification With Belief Decision Trees. Artificial Intelligence: Methodology, Systems, and Applications. (ss.80-90) Berlin: Springer

Faliagka, E., Iliadis, L., Karydis, I., Rigou, M., Sioutas, S., Tsakalidis, A. ve Tzimas G. (2014) On-line Consistent Ranking On E-recruitment: Seeking The Truth Behind A Well-Formed CV. Artificial Intelligence Review. 42: 515–528.

Fogg, A. (05.07.2017). A History of Machine Learning and Deep Learning. <https://www.import.io/post/history-of-deep-learning/>, (05.12.2020).

Foote, K.D. (07.02.2017). A Brief History of Deep Learning. <https://www.dataversity.net/brief-history-deep-learning/>, (05.12.2020).

Hastie, T., Tibshirani, R. ve Friedman, J. (2009). The Elements of Statistical Learning. New York: Springer.

James, G., Witten, D., Hastie, T. ve Tibshirani, R. (2013). An Introduction To Statistical Learning. New York: Springer.

Joshi, N. (9 Şubat 2019). Recruitment Chatbots: Is The Hype Worth It? <https://www.forbes.com/sites/cognitiveworld/2019/02/09/recruitment-chatbot-is-the-hype-worth-it/?sh=7eda771d4083>, (26.12.2020).

Kavzaoğlu, T. ve Çölkesen, İ. (2010) Destek Vektör Makineleri ile Uydu Görüntülerinin Sınıflandırılmasında Kernel Fonksiyonlarının Etkilerinin İncelenmesi. Harita Dergisi. (144): 73-82.

Khanna, S.K. (2019). Machine Learning V/S Deep Learning. International Research Journal of Engineering and Technology. 6(2): 455-458.

Kılınç, D., Borandağ, E., Yücalar, F., Özçift, A. ve Bozyiğit, F. (2015). Yazılım Hata Kestiriminde Kolektif Sınıflandırma Modellerinin Etkisi. IX.Ulusal Yazılım Mühendisliği Sempozyumu Kitabı (ss.113-121). Düzenleyen Yaşar Üniversitesi. İzmir. 9-11 Eylül 2015.

Kılınç, M. , Tarhan, Ç. & Aydın, C. (2020). Could Mobile Applications' Success be Increased via Machine Learning and Business Intelligence Methods? . Avrupa Bilim ve Teknoloji Dergisi , (20) , 805-814 . DOI: 10.31590/ejosat.793069.

Kocakafa, T. (26 Şubat 2020). Aşırı Öğrenme (Overfitting). <https://www.veribilimiokulu.com/blog/overfitting/>, (27.12.2020).

Leong, C. (2018). Technology & Recruiting 101: How It Works and Where It's Going. *Strategic HR Review*. 17(1): 50-52.

LinkedIn, (2020). LinkedIn Recruiter Datasheet. <https://business.linkedin.com/talent-solutions/recruiter/recruiter-datasheet>, (10.12.2020).

LinkedIn, (27 Mart 2018). The Rise of Analytics in HR. [https://business.linkedin.com/content/dam/me/business/en-us/talent-solutions/talent-intelligence/workforce/pdfs/Final\\_v2\\_NAMER\\_Rise-of-Analytics-Report.pdf](https://business.linkedin.com/content/dam/me/business/en-us/talent-solutions/talent-intelligence/workforce/pdfs/Final_v2_NAMER_Rise-of-Analytics-Report.pdf), (10.12.2020).

Lipton, Z.C., Elkan, C. ve Narayanaswamy, B. (2014). Optimal Thresholding Classifiers To Maximize F1 Measure. *Machine Learning and Knowledge Discovery In Databases Part II*. Berlin: Springer.

Marr, B. (14 Aralık 2008). The Amazing Ways How Unilever Uses Artificial Intelligence To Recruit & Train Thousands Of Employees. <https://bernardmarr.com/default.asp?contentID=1766>, (20.12.2020).

McKinney, W. ve Pandas Geliştirme Ekibi. (30 Ekim 2020). Pandas: Powerful Python Data Analysis Toolkit. <https://pandas.pydata.org/docs/pandas.pdf>, (15.11.2020).

Nizam, H. ve AKIN, S.S. (2014). Sosyal Medyada Makine Öğrenmesi ile Duygu Analizinde Dengeli ve Dengesiz Veri Setlerinin Performanslarının Karşılaştırılması. 19.Türkiye'de İnternet Konferansı Bildiriler Kitabı (ss.129-136), Düzenleyen Yaşar Üniversitesi. İzmir. 27-29 Kasım 2014.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M. ve Duchesnay, E. (2011). Scikit-learn: Machine Learning in Python. *Journal Of Machine Learning Research*. 12(85): 2825–2830.

Polat, S. (2017). Yazılım Hata Kayıtlarının Makine Öğrenmesi Yöntemleriyle Kümelenerek, Hataya Sebep Olan Bileşenlerin Tespit Edilmesi. 11.Ulusal Yazılım Mühendisliği Sempozyumu Kitabı (ss. 444-453), Düzenleyen Atılım ve Alanya Hamdullah Emin Paşa Üniversitesi. Alanya. 18-20 Ekim 2017.

Polikar, R. (2012). Ensemble Learning. *Ensemble Machine Learning*. (ss.1-34). Boston: Springer.

Portilla J. (2020). Python for Data Science and Machine Learning Bootcamp. <https://www.udemy.com/course/python-for-data-science-and-machine-learning-bootcamp/>, (20.12.2020).

Samuel, A.L. (1959). Some Studies in Machine Learning Using The Game of Checkers. *IBM Journal of Research and Development*. 3(3): 211-229.

Sathya, R. ve Abraham, A. (2013). Comparison of Supervised and Unsupervised Learning Algorithms for Pattern Classification. *International Journal of Advanced Research in Artificial Intelligence*. 2(2): 34-38.

Scikit-learn, (2020a). Nearest Neighbors Classification, [https://scikit-learn.org/stable/auto\\_examples/neighbors/plot\\_classification.html#sphx-glr-auto-examples-neighbors-plot-classification-py](https://scikit-learn.org/stable/auto_examples/neighbors/plot_classification.html#sphx-glr-auto-examples-neighbors-plot-classification-py), (20.12.2020).

Sokolova, M. ve Lapalme, G. (2009). A Systematic Analysis Of Performance Measures For Classification Tasks. *Information Processing & Management*. 45(5):427-437.

Steinwart, I. ve Christmann, A. (2008) *Support Vector Machines*. New York: Springer.

Streamlit, (2020). Welcome To Streamlit. <https://docs.streamlit.io/en/stable/>, (20.12.2020).

Strohmeier, S. ve Piazza, F. (2015). Artificial Intelligence Techniques In Human Resource Management-A Conceptual Exploration. *Intelligent Techniques in Engineering Management*. (ss.149-169) Cham: Springer.

Sutton, R.S. ve Barto, A.G. (2018). *Reinforcement Learning: An Introduction*. Cambridge: The MIT Press.

Şeker, A., Diri, B. ve Balık, H.H. (2017). Derin Öğrenme Yöntemleri ve Uygulamaları Hakkında Bir İnceleme. *Gazi Mühendislik Bilimleri Dergisi*. 3(3): 47-64.

Turing A.M. (1950). Computing Machinery and Intelligence. Mind. 59(236): 433-460.

Türkiye Büyük Millet Meclisi, (07 Nisan 2016). Kişisel Verilerin Korunması Kanunu. Resmi Gazete. <https://www.resmigazete.gov.tr/eskiler/2016/04/20160407.pdf>, (20.11.2020).