



Yapay Öğrenme Teknikleriyle Milli Eğitim Bakanlığı'nda Yarı Zamanlı Öğretmenlerin Görevlendirilmesi

Assignment of Part-Time Teachers in the Ministry of National Education with Artificial Learning Techniques

Ertürk ERDAĞI

Türk Telekom Şehit Murat Naiboğlu Anadolu Lisesi
İstanbul, Türkiye
erturkerdagi@gmail.com
ORCID: 0000-0001-8619-8879

Volkan TUNALI

Maltepe Üniversitesi
Yazılım Mühendisliği (İngilizce) Bölümü
İstanbul, Türkiye
volkan.tunali@gmail.com
ORCID:0000-0002-2735-7996

Öz

Bu çalışmada okullarda görev alacak yarı zamanlı öğretmenlere yönelik verinin çeşitli sınıflandırma algoritmaları kullanılarak, yapılan atamaların başarı ölçümünü gözlemlemek ve yapılan işlemin yapay zeka destekli sınıflandırma algoritmaları ile yürütülmesi durumuna yönelik analizi gerçekleştirilmiştir. Çalışmada web tabanlı bir uygulama üzerinden yapılan başvurular ile elde edilen verilerin, gereksinimler doğrultusunda yapılan atama ölçütlerine yönelik başarıyı ölçülmüştür. Bunun için 3053 adayın başvurduğu sistemden gereksinim doğrultusundaki 894 adayın atanmasına yönelik öznelikler üzerinde çalışılmıştır. Altı farklı sınıflandırıcı üzerinde yapılan çalışmada 0,71 Doğruluk ve 0,77 F-ölçüsü ile en iyi sonucu Rassal Orman Sınıflandırıcısı vermiştir. Yapılan bu çalışma ile atama ölçünlerinin ağırlıklarına göre bu çalışmanın uygun sınıflandırıcılar ile yürütülmesi ve bu alanda kullanılabileceği gösterilmiştir.

Anahtar sözcükler: Milli Eğitim Bakanlığı, Öğretmen, Sınıflandırma, Yapay Öğrenme

Abstract

In this study, the data for part-time teachers who will work in schools were analyzed using various classification algorithms to observe the success measurement of the assignments and automate the process with artificial intelligence-supported

classification algorithms. In the study, the performance of the data obtained through the applications made through a web-based application, for the assignment criteria made in line with the needs, was measured. For this, the attributes for the appointment of 894 candidates in line with the needs from the system applied by 3053 candidates were studied. In the study conducted on six different classifiers, the Random Forest Classifier gave the best result with an accuracy value of 0.71 and an f-score of 0.77. With this study, it has been shown that this study can be automated with appropriate classifiers and used in this field according to the weights of the assignment criteria.

Keywords: Ministry of National Education, Teacher, Classification, Artificial Learning

1. Giriş

Milli Eğitim Bakanlığı'na bağlı kurumlarda görev yapan öğretmenlerin sayısının yeterli olmadığı durumda İlçe Milli Eğitim Müdürlükleri tarafından gereksinim doğrultusunda yerel bazda Yarı Zamanlı Öğretmenler görevlendirilir. Görevlendirilecek öğretmenler çeşitli ölçütler doğrultusunda seçilir. Bu ölçütler arasında gereksinim duyulan alanda mezun olunması, pedagojik formasyon özelliği, daha önce görev yapıp yapmadığı gibi birçok nitelik bulunmaktadır. Bu durum eğitim planlamasının yalnızca nicel değil, aynı zamanda nitel bir olgu olduğunu kanıtlar niteliktedir [1]. Her eğitim öğretim yılı başında yapılan bu işlem adayların internet üzerinden bilgilerini verdikleri ve ardından e-devlet üzerinden yaptıkları başvurular ile gerçekleşmektedir. Sistem ASP.NET yapısında

ve C# dili kullanılarak hazırlanan bir web uygulaması ile verilerin SQL Server üzerindeki bir veri tabanında saklanması üzerine hazırlanmıştır. Adayların belirlenmesi sırasında kullanılacak öznitelikler çoğu zaman puanlama yolu ile belirlenirken bazı durumlarda farklı öznitelikler ile ya da gereksinim belirtilen alanda öğretmen olmaması nedeniyle puan dışına çıkmaktadır. Bu noktada kararlı düzeyde bir analiz ve seçim yöntemi için sınıflandırma algoritmaları kullanılmış ve elde edilen sonuçlar karşılaştırılarak yeni gelecek bir veriye ilişkin doğru analiz yöntemi ile karar verilebilmesi amaçlanmıştır. 3053 adayın çeşitli özniteliklere ilişkin puanlama yapılması ile 894 aday seçilme işlemi için bu durum daha kararlı ve yapılacak sınıflandırma yöntemleri sonrasında belirlenmesi daha sağlıklı olacaktır. Yapılan çalışma yapay zekanın alt kolu olan makine öğrenmesine dayalı algoritmaların kullanımı şeklinde gerçekleşmektedir. Makine öğrenmesi içerisinde yer alan farklı algoritmaların kullanımı veri için özel ve hassas değerler üretilmesini sağlamaktadır[2]. Bu çalışma ile zaman ve maliyet konularında tasarruf sağlanması ve yapılacak çalışmanın daha iyi sonuç vermesine yönelik kazanımlar elde edilebilmektedir. Çalışma kapsamında veri kümesi üzerinde yapay öğrenme destekli bir sınıflandırma çalışması için altı farklı algoritma kullanılmıştır.

Yarı zamanlı öğretmenlerin istihdamı gibi önemli bir hususun hem zaman hem de maliyet yönünden doğru bir yöntemin bulunmasının yanında maddi hataların önüne de geçilmesini sağlayacak bir yöntem önerilmektedir.

2. Literatür Taraması

Sınıflandırma çalışmaları kapsamında öneri sistemleri geniş kullanım alanına sahiptir. Bu bağlamda sayısal bir öge, kişi ya da karar kısımlarında sınıflandırıcı algoritmalar kullanılmaktadır. Çalışma kapsamında çeşitli bilgilerden elde edilerek nihayetinde elde edilen sonuç öneri kapsamında değerlendirilebilmektedir. Özgöbek ve ark. içerik tabanlı filtreleme çalışması kapsamında daha önce var olan bilgiler üzerinden öğelerin benzerliklerini kullanarak bir çalışma yürütmüştür [3]. Sayar ve Turdaliev makine öğrenmesi yöntemleri kullanarak kullanıcılar için otel öneri sistemi konusunda çalışma yürütmüşlerdir. Müşterilerin memnuniyet durumları ve fiyat endeksi kullanılarak makine öğrenmesi ile otel öneri sistemi oluşturulmuştur. Çalışma kapsamında eğitim için otellerin birbirlerine göre üstün yanları belirlenmiş, yeni verilere ilişkin diğer oteller ile karşılaştırılması sonucunda daha iyi bir fırsat olup olmadığına karar verilmesini sağlayan bir yapı kurulmuştur [4]. Dünder ve Kakişim ürünlerin özellikleri arasındaki güçlü korelasyon durumunu kullanarak kıyafet öneri sistemi üzerine çalışma gerçekleştirmişlerdir. Çalışma kapsamında ağırlıklı graf yapısı üzerinde ürünlere ait özelliklerin ağırlıklandırılması ve bunun öğrenilmesine dayanan sistem sayesinde müşterilerin alışveriş davranışlarına göre en uygun kıyafetin önerilmesi sağlanmıştır[5]. Bal ve ark. işe alım süreçlerinde yapay zekanın kullanımına ilişkin yaptığı çalışmada karar vericilerin işlerini kolaylaştırmada ve özellikle nesnel olmayan kararların önüne geçilmesi anlamında yapay zekanın olumlu sonuçlar verdiğini belirtmişlerdir. İşe alım süreçlerinde kullanılan "Başvuru İzleme Sistemi" adayın özgeçmiş bilgileri ile iş tanımı arasında doğru bir eşleştirme yapılmasını ve sürecin sağlıklı

yürütülmesini sağlamaktadır [6]. Gültekin ve ark. aday değerlendirme aşaması için Çok Kriterli Karar Verme yöntemi kapsamında en iyi kararın verilmesi için bulanık mantık kullanmışlardır. Karar verme sürecinin yorumlara dayalı değişken bir süreç şeklinde ilerlemesi ve içerisinde belirsizliklerin bulunması bu çalışmada bulanık mantığın kullanımında önemli bir etken olmuştur [7].

3. Yöntem

3.1 Veri Kümesi

Veri kümesi web uygulaması üzerinden başvuran adayların bilgilerinin kaydedildiği SQL Server veri tabanı içerisinde birkaç tablo kullanılarak elde edilmiştir. Elde edilen veriler toplamda 3053 adayın bulunduğu veriler içerisinde kişisel bilgisine yönelik çalışma yapılmamış, tüm kayıtlar anonim olarak toplanmıştır. Toplam 3053 adayın başvuru yaptığı dönemde 894 öğretmen gereksinimi belirlenmiştir. Bu çalışmada yapılan seçim işlemine ilişkin veriler kullanılarak çeşitli sınıflandırma algoritmaları kullanılarak karşılaştırma ve analiz yöntemleri tartışılmıştır. 894 adayın seçileceği ve 3053 adayın başvurduğu sistemde seçim kriterini oluşturan birçok öznitelik bulunmaktadır.

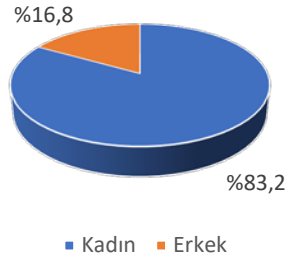
Çizelge-1: Veri seti içerisindeki öznitelikler

Öznitelik Adı	Öznitelik Açıklaması
CandidateID	İlgili adaya ilişkin nümerik formatta benzersiz bir değer içeren ID değeridir.
FirstName	Adaya ilişkin isim bilgisini saklar.
LastName	Adaya ilişkin soyad bilgisini saklar.
GenderText	Adaya ilişkin cinsiyet bilgisini saklar.
LastSchoolName	Daha önce görev yapmışsa son görev yaptığı okulun ismini saklar.
GraduationBranchName	Adayın mezun olduğu bölüm bilgisini saklar.
GraduationTypeName	Adayın mezuniyet türünü (lisans, önlisans, yüksek lisans vb.) bilgisini saklar.
LastTerm	Adayın son görev yaptığı dönemi saklar. Eğer ilk kez görev yapacaksa bu alan boş kalır.
DistrictName	Adayın ikamet ettiği ilçe adını saklar.
BranchRequest	Adayın görev yapmak istediği branş adını saklar.
MissionTime	Adayın daha önce ne kadar süre görev yaptığı bilgisini nümerik olarak saklar.
RequestSchoolName	Adayı görev yapmak istediği okulun ismini saklar.
IsAcceptOutOfPref	Adayın tercihi dışında bir okul ya da branşta görev yapmak isteyip istemediğine ilişkin veriyi saklar. Tercih dışında okul ya da branşı kabul eden adaylar için bu alan 1, etmeyenler için 0 olacak şekilde nümerik veri saklanır.
IsHasSertificate	Adayın özel branşlarda (Özel Eğitim Öğretmenliği, İşitme Engelliler Öğretmeni vb.) gerekli sertifikasının olup olmadığı bilgisini saklar. Bu durum sertifikası olanlar için 1, olmayanlar için 0 değerini saklar.
IsInBlackList	Adayın önceki yıllarda herhangi bir nedenle görevden çıkarılması ya da ilgili yöneticisi tarafından olumsuz görüş

	belirtilmesi neticesinde kara listeye alınıp alınmadığını gösterir. Kara listedeki adaylar için bu alan 1, diğer adaylar için 0 değerini saklar.
CandidateStatus	Çalışmanın nihai noktasını oluşturan bölümdür. İlgili adayın durumunu belirten veriyi saklar. Aday için mülakata girebilir ya da hak kazanamadı şeklinde olmak üzere iki farklı durumu kategorik olarak saklar.

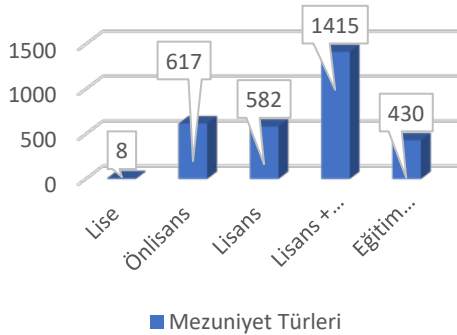
Öznelikler karar verme sürecinde sağlıklı sonuç elde edebilmek adına başvuru sırasında istenilen bilgilerden elde edilmiştir. Öznelik belirleme sürecinde aday için karar verme sürecinde etkili olabilecek bilgiler kullanılmıştır. Adayın deneyim süresi, branşı vb. bilgiler ihtiyaç duyulacak branş ve sayı için kritik önem arz etmektedir. Veri seti içerisindeki 16 öznelik açıklamasıyla birlikte Çizelge-1’de sunulmuştur.

Şekil-1’de yer alan adayların cinsiyete göre dağılım grafiği incelendiğinde başvuran adayların %83,2’sinin kadınlardan oluştuğu, %16,8’inin erkeklerden oluştuğu gözlemlenmektedir.

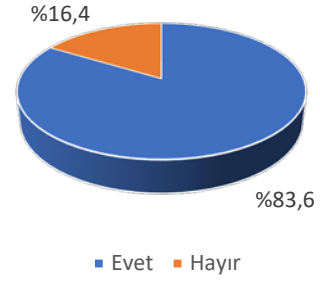


Şekil-1: Adayların cinsiyete göre dağılımlarına ilişkin grafik

Şekil-2’de yer alan beş farklı mezuniyet tipinde başvuran adayların dağılımına ilişkin grafik incelendiğinde en fazla mezuniyet türünün 415 aday ile Lisans + Formasyon türünde olduğu gözlemlenmektedir. Bu mezuniyet tipini 617 aday ile önlisans, 582 ile lisans, 430 ile eğitim fakültesi ve 8 aday ile lise mezuniyet türü takip etmektedir.

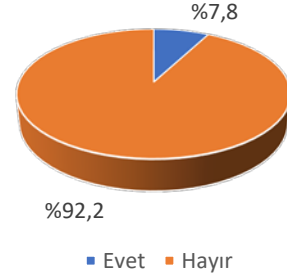


Şekil-2: Adayların mezuniyet türlerine göre dağılımlarına ilişkin grafik



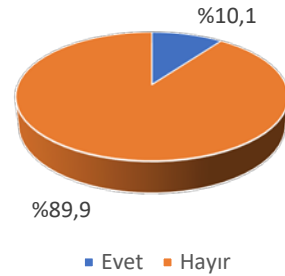
Şekil-3: Adayların tercihleri dışında görevleri kabul edip etmeme durumlarının ilişkin grafik

Şekil-3’teki grafik incelendiğinde başvuruda bulunan adayların %83,6’sının tercihleri dışında bir göreve atanmayı kabul ettiği, %16,4’ünün ise bu seçeneği kabul etmediği gözlemlenmektedir. Bu kısımdaki tercih dışı bilgisi adayın seçtiği okulların dışında bir okulda görev yapıp yapmaması ya da seçmiş olduğu branş dışında başka bir branşta da görev yapıp yapamayacağı konusunu içermektedir.



Şekil-4: Adayların özel öğretim branşında gerekli sertifika sahibi durumuna ilişkin grafik

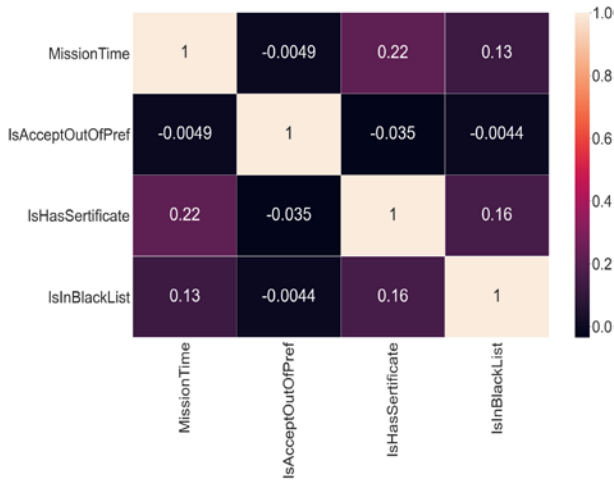
Özel öğretim branşında kullanılmak için aranan sertifika, Şekil-4’te görüldüğü gibi başvuran adayların %7,8’lik bölümünde bulunmaktadır. Başvuran adayların %92,2’sinde sertifika olmadığı gözlemlenmektedir. Burada sertifika bilgisi özel öğretim branşları için öncelikli aday belirlemede kullanılmaktadır.



Şekil-5: Adayların kara liste durumlarına ilişkin grafik

Şekil-5’te bulunan grafik incelendiğinde başvuruda bulunan adayların %10,1’lik bir kısmının kara listede olduğu görülmektedir. Kara liste daha önce görev yaparken çeşitli sebepler ile eğitim öğretim ortamına uygun olmayan, resmi yollar ile bunların belgelendirildiği, idari ve adli soruşturmalar sonucunda eğitim öğretim ortamına uygun olmadığı gözlemlenen ve süreç başında olumsuzluğun tekrar

yaşanmaması adına sistem yöneticisi tarafından yapılan bir müdahaleyi içeren bir bilgiyi saklamaktadır.



Şekil-6: Korelasyon matrisi grafiği

Korelasyon matrisi değişkenlerin birbiriyle olan ilişki yönünü ve gücünü belirler [8]. Korelasyon matrisinde her biri ikili vektör ilişkisi için -1 ile 1 arasında değer verilmektedir. Bu değer 1'e yakınsa korelasyonuna bakılan iki veri vektörü arasında güçlü bir doğru orantı, -1'e yakınsa güçlü bir ters orantı vardır denilir. Eğer korelasyon değeri 0'a yakın ise veriler arasında lineer bir ilişkiden söz edilemez. Şekil-6'da üzerinde çalışma kapsamında oluşturulan korelasyon matrisinde değişkenlerin birbirleri ile çok fazla ilişkisi olmadığı, en yüksek ilişkinin görev süresi verisine ait MissionTime özneliği ile özel branşlar için sertifika bilgisini saklayan IsHasCertificate özneliği olduğu 0,22 gibi küçük bir ilişki değeri ile gözlemlenmektedir. Korelasyon matrisi incelendiğinde özneliklerin birbirleri ile lineer bir ilişki içinde olmadığı görülmektedir. Siyah renkte belirtilen bu öznelik çiftlerinin değerleri 0'a yakın durumdadır.

3.2 Ön İşleme Aşaması

Ön işleme işlemine veri seti içerisinde eksik bilgi olup olmadığı ve bunun genel veri içerisindeki yüzdesini gösteren bir fonksiyonun yazılması ile başlanmıştır. Fonksiyonun sonuçları Çizelge-2'de sunulmuştur.

Fonksiyon sonucunda LastSchoolName ve LastTerm özneliklerinde boşluklar olduğu tespit edilmiştir. Bu iki öznelik adayın daha önce görev yaptığı okul ve dönem bilgilerini saklamaktadır. Daha önce görev yapmayan bir aday için bu kısımlar boş olduğundan, %66'yı aşan ve büyük bir oranda boş veri içerdiğinden sınıflandırma aşamasında katkı sunmayacağı için veri seti üzerinden kaldırılmıştır. Ön işleme öncesinde 16 öznelik ve 3053 satır içeren data seti içerisinde sınıflandırma işlemi içerisinde önemi olmayacak durumda olan ve analizde özellikle sayısal değerinden dolayı farklı etki oluşturacak CandidateID özneliği, ad ve soyad bilgisini içeren FirstName ve LastName öznelikleri ön işleme işlemleri kapsamında veri seti üzerinden kaldırılmıştır. Adayın cinsiyet bilgisini içeren GenderText, mezun olunan bölüm bilgisini içeren GraduationBranchName, ikamet edilen ilçe bilgisini saklayan DistrictName, tercih edilen branş bilgisini saklayan

BranchRequest, tercih edilen okul bilgisini saklayan RequestSchoolName öznelikleri için one-hot encoding [9] işlemi uygulanmıştır. Bu yöntem makine öğrenmesi çalışmalarında kategorik verilerin sayısal değerlere dönüşmesinde kullanılmaktadır. Verilerin temsil edilmesinde kolaylık sağlayan bu yöntemde önce kategorik değerler tamsayı değerleriyle eşleştirilir. Sonrasında her bir tamsayı değeri 1 ile işaretlenmiş tamsayı indeksi dışındaki tüm değerleri sıfır olan bir ikili vektör olarak temsil edilir. Elde edilen sonuçlar "encode" edilmiş verileri temsil ederek makine öğrenmesi algoritmalarında veri olarak kullanılmaktadır. Sınıflandırma işleminin sonucu olan ve çalışmanın nihai noktasını gösteren CandidateStatus özneliğinde bulunan iki ihtimal değeri ilgili öznelik içerisinde nümerik değerlere çevrilmiştir. Bu noktada olumlu sonuç olan ve kabul edilen adaylar için 1, diğer adaylar için 0 değeri olacak şekilde güncelleme yapılmıştır. Daha önce yapılan görev süresi verisini içeren MissionTime verisi üzerinde ölçeklendirme işlemi yapılmıştır. MinMaxScaler kullanılarak yapılan bu işlem sonrasında ilgili öznelikte en yüksek verinin 1, en düşük verinin 0 olacağı şekilde dağıtım yapılmıştır.

Çizelge-2: Veri kümesi içerisindeki eksik değerler ve genel veri üzerindeki yüzdesi

Öznelik	Eksik Değerler	% Değeri
CandidateID	0	0,00
FirstName	0	0,00
LastName	0	0,00
GenderText	0	0,00
LastSchoolName	2015	66,02
GraduationBranchName	0	0,00
GraduationTypeName	0	0,00
LastTerm	2018	66,12
DistrictName	0	0,00
BranchRequest	0	0,00
MissionTime	0	0,00
RequestSchoolName	0	0,00
IsAcceptOutOfPref	0	0,00
IsHasCertificate	0	0,00
IsInBlackList	0	0,00
CandidateStatus	0	0,00
IsAcceptOutOfText	0	0,00
IsCertificateText	0	0,00
IsInBlackListText	0	0,00

3.2 Kullanılan Sınıflandırma Algoritmaları

Bu çalışma kapsamında altı adet sınıflandırıcı kullanılmıştır. Farklı sınıflandırıcıların kullanım amacı makine öğrenmesinin temelinde olan veriye özgü farklı algoritmalar arasında beklenen sonuç için en iyi sonucun bulunmasını sağlamak ve bu sayede zaman tasarrufu sağlamaktır.

Kullanılan algoritmalarından ilki Bayes teoremine [10] dayanan Naive Bayes sınıflandırıcısıdır. Dengesiz veri kümelerinde iyi sonuçlar verebilen bir sınıflandırıcı olması sebebiyle bu

çalışma içerisinde tercih edilmiştir. Çünkü veri seti içerisinde farklı özniteliklerin birbirleriyle korelasyon ilişkisi düşük olup özellikle bazı niteliklerde aşırı yüklenmenin bir tarafa olduğu gözlemlenmektedir. Gaussian Naive Bayes (GaussianNB) sınıfı üzerinden yapılan bu işlem içerisindeki üç adet sınıftan Bernoulli Naive Bayes (BernoulliNB) tercih edilmiştir. BernoulliNB sınıfı sonuç veri kümesinin ikili olmasını bekleyen bir sınıf olduğundan ve sonuç kısmı adayın mülakata giremez ya da mülakata hak kazandı şeklinde ikili bir veri yapısı içerdiğinden kullanımı daha fazla avantaj sağlayacaktır.

Çalışma kapsamında kullanılan diğer algoritma Karar Ağacı Sınıflandırıcısı'dır. Veri seti üzerinde çok fazla özneliğin olması ve özniteliklerin farklı durumlarına göre bölünerek çalışma ve yinelemeli model üzerinden sonuca ulaşması, ayrıca uygulama ve hız bakımından avantaj getirmesi nedeniyle bu algoritma, çalışmaya dahil edilmiştir.

Kullanılan üçüncü algoritma K-En Yakın Komşu Sınıflandırıcısı'dır. Gürültüye karşı dirençli bir algoritma olması sebebiyle bu çalışmada kullanılmıştır. Veri kümesi içerisinde kara liste ve görev süresi gibi parametrik özniteliklerin gürültü oluşturabileceği ve diğer özniteliklerin sınıflandırmaya olumlu katkı yapacağı noktalarda, bu ve bunun gibi öznitelikler gürültü oluşturabilecek durumda olacağından, çalışma kapsamında avantaj oluşturabilecek noktada kullanılmıştır.

Çalışmanın dördüncü algoritması Rassal Orman Sınıflandırıcısı'dır [11]. Bu algoritmanın seçilmesinin nedeni algoritmanın özniteliklerin önem derecesini gösterebilme yeteneğidir. Buradaki önem durumu o özneliğin bağımlı değişkendeki varyansın açıklanmasına sunduğu katkı durumudur. Dolayısıyla yapılacak çalışmada sonucu belirleyecek öznitelikler belirlenebilir, katkı sunmayan ya da olumsuz yönde etkide bulunan öznitelikler devre dışı bırakılabilir. Veri setimiz içerisinde farklı özniteliklerin sonuca yapacağı katkı tam olarak tahmin edilemediğinden farklı sonuçlar doğuran ya da etkisi olmayan öznitelikler bu algoritma ile devre dışı bırakılmakta, bu sayede hem performans hem de sonuç anlamında olumlu kazanımlar elde edilmiş olur.

Kullanılan beşinci algoritma Karar Destek Sınıflandırıcısı'dır. Veri seti içerisinde doğrusal ve doğrusal olmayan öznitelikler bulunduğundan bu işlem öznelik seçimi ve ön işleme sonrasında elde edilen yeni nümerik özniteliklere göre değişiklik gösterebileceği düşünüldüğünden çalışma içerisine dahil edilmiştir.

Kullanılan son algoritma ise Yapay Sinir Ağı Sınıflandırıcısı'dır. Veri seti içerisinde sonuca etki eden güçlü özneliğin belirlenmesi sürecinde farklı bileşenlerden tek bir çıktı elde edilebilmesini sağladığından bu algoritma kullanılmıştır. Modelin eğitimi için işlemin birkaç kez tekrar edilebilmesi ve en güçlü sonuca ulaşabilmesini sağlamak bu algoritmanın güçlü yönünü oluşturmaktadır.

4. Deneysel Sonuçları ve Tartışma

Tüm sınıflandırıcıların çalışma öncesinde veri üzerinde ön işleme ve normalizasyon işlemi gerçekleştirilmiştir. Çalışma kapsamında k-katlamalı çapraz doğrulama yöntemi kullanılmıştır. Bu yöntemde eğitim kümesi k değerine göre eşit alt kümelerle bölünür. Her yineleme için k alt kümelerinden biri test, k-1 alt kümeleri eğitim verisi olarak kullanılmaktadır. k kez tekrar eden bu işlem sonucunda ortalama hata değeri hesaplanır [12]. Tüm sınıflandırıcılar için sonuçlar k-katlamalı çapraz doğrulama (k=10) kullanılarak elde edilmiştir. Bu kısımda 10 değeri literatürde en fazla kullanılan değer olduğundan tercih edilmiştir [13]. Bu yöntem modelin değerlendirilmesinde önemli olup, aşırı öğrenme ve eksik öğrenme gibi sorunların önüne geçmeyi sağlamaktadır [14]. Aşırı öğrenme problemi eğitim kümesindeki verileri çok iyi öğrenirken, modele yeni gelen veriler üzerinde doğru tahmin gerçekleştirememeye neden olur. Eksik öğrenme ise eğitim esnasında doğru bir şekilde öğrenimin gerçekleşmemesi sonucu yeni verilere karşı doğru tahminleme işleminin gerçekleşmemesidir. Bu iki sorun doğru sınıflandırma algoritması ile çözülebilmektedir. Sınıflandırma algoritmalarının çalışması için düzenlenen parametreler Çizelge-3'te sunulmuştur. Çalışma sonucunun değerlendirilmesinde dört farklı metrik kullanılmıştır. Bunlardan ilki olan tutturma pozitif olarak tahmin edilen değerlerin kaç tanesinin pozitif olduğuna ilişkin bir ölçüttür. İkinci metrik olan bulma doğru tespit edilen pozitif sınıfların tüm pozitiflere oranı ile hesaplanmaktadır. Üçüncü metrik olan f-skor değeri tutturma ve bulma değerlerinin harmonik ortalaması ile bulunmaktadır. Burada aritmetik ortalama yerine harmonik ortalama kullanılmasının nedeni aykırı nitelikteki durumların göz ardı edilmemesi gerekliliğindedir. Dördüncü metrik olan doğruluk, yapılan doğru tahminin tüm veriye ilişkin oranının elde edilmesini sağlar. Tek başına kullanımı dengesiz bir veri seti için hatalı sonuçlar verebilmektedir.

Öğrenim Durumu	Mezun Olduğu Bölüm	Durum	Okul	Branş	Kaydet
Yüksek Lisans	TÜRK DİLİ VE EDEBİYATI BÖLÜMÜ	Atama bekleniyor	---Seçiniz---	---Seçiniz---	Kaydet
Lisans	SOSYAL BİLGİLER EĞİTİMİ BÖLÜMÜ	Atama bekleniyor	---Seçiniz---	---Seçiniz---	Kaydet
Lisans	İLAHİYAT BÖLÜMÜ	Atama geçici olarak yapıldı	Mehmet Corcor İlkokulu	Sınıf Öğretmenliği	Kaydet
Lisans	İLAHİYAT BÖLÜMÜ	Atama bekleniyor	---Seçiniz---	---Seçiniz---	Kaydet
Lisans	İLAHİYAT BÖLÜMÜ	Atama geçici olarak yapıldı	Türhan Feyzioglu Mesleki ve Teknik Anadolu L	Din Kültürü Ve Ahlak Bilgisi Öğretmeni	Kaydet
Lisans	İLAHİYAT BÖLÜMÜ	Atama geçici olarak yapıldı	Ertugrul Gazi Ortaokulu	Din Kültürü Ve Ahlak Bilgisi Öğretmeni	Kaydet

Şekil-8: Sonuç kontrol ekranı

Çizelge-3: Sınıflandırma algoritmaları parametreleri

Sınıflandırıcı	Parametreler
Karar Ağacı	cv=10, scoring=['f1_weighted', 'precision', 'recall', 'accuracy', 'roc_auc']
K-En Yakın Komşu	cv=10, scoring=['f1_weighted', 'precision', 'recall', 'accuracy', 'roc_auc']
Naïve Bayes	cv=10, scoring=['f1_weighted', 'precision', 'recall', 'accuracy', 'roc_auc']
Yapay Sinir Ağı	activation='logistic', solver='adam', max_iter=2000, random_state=42, hidden_layer_sizes=(4,)) cv=10, scoring=['f1_weighted', 'precision', 'recall', 'accuracy', 'roc_auc']
Karar Destek	cv=10, scoring=['f1_weighted', 'precision', 'recall', 'accuracy', 'roc_auc']
Rassal Orman	cv=10, scoring=['f1_weighted', 'precision', 'recall', 'accuracy', 'roc_auc']

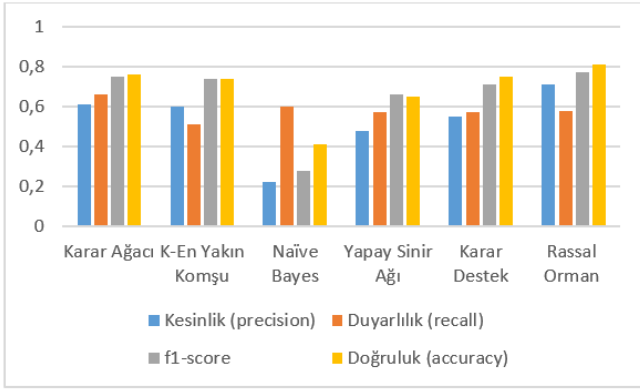
Sınıflandırıcının çalışmasına etki eden komşu değerinin değiştirilmesi ile birlikte sınıflandırıcının performansında çok küçük değişiklikler görülmüştür. Bu noktada ilgili parametre 3 iken 5 değerine yükseltilmiş ve 0,72 olan doğruluk değeri 0,74'e yükselmiştir. İlgili parametrenin diğer tek sayılarda daha düşük bir oranda başarı sağladığı görüldüğünden optimum değer olarak 5 değerinde karar kılınmıştır. Nihai sonuç Çizelge-4'te görülmektedir.

Çizelge-4: Altı sınıflandırıcı ile elde edilen sonuçlar

Sınıflandırıcı	Kesinlik (precision)	Duyarlılık (recall)	F-skor (f1-score)	Doğruluk (accuracy)
Karar Ağacı	0,61	0,66	0,75	0,76
K-En Yakın Komşu	0,60	0,51	0,74	0,74
Naïve Bayes	0,22	0,60	0,28	0,41
Yapay Sinir Ağı	0,48	0,57	0,66	0,65
Karar Destek	0,55	0,57	0,71	0,75
Rassal Orman	0,71	0,58	0,77	0,81

Çalışmaya dair sonuçlar ayrıca görselleştirilmiş ve adayların isimleri olmaksızın sonuçları hem okul hem de branş bazında listelenmiştir. Uygulamanın bu kısmı web üzerinden kontrol edilebilecek şekilde sonuçlandırılmıştır. Belirtilen bu husus Şekil-8'de görsel olarak eklenmiştir. Bu kısımda turkuaz renkteki adaylar ataması kriterler doğrultusunda gerçekleşenleri, beyaz renktekiler ise henüz bir ataması yapılmayan adayları temsil etmektedir.

Adaylar ile elde edilen sonuçlar görsel olarak kriterlerin kontrolünde ve adaylar ile ilgili toplu liste alımı için web uygulamasıyla da desteklenmiştir.



Şekil-9:Altı sınıflandırıcı algoritma ile elde edilen sonuçlara ait grafik

5. Sonuç

ASP.NET yapısında ve C# dili kullanılarak hazırlanan bir web uygulaması ile adayların başvuruda bulunduğu ve verilerin SQL Server üzerindeki bir veri tabanında saklanmasıyla gerçekleşen uygulama içerisinde elde edilen veri seti adayların kişisel bilgileri dışındaki özellikler ile elde edilmiştir. Toplamda 3053 aday ve ihtiyaç dahilinde görevlendirilecek 894 öğretmene ait işaretli veri elde edilmiştir. Bu çalışma ile zaman ve maliyet konularında tasarruf sağlanması ve yapılacak çalışmanın daha iyi sonuç vermesine yönelik kazanımlar elde edilebilmektedir. Çalışma kapsamında veri seti üzerinde yapay öğrenme destekli bir sınıflandırma çalışması için altı farklı algoritma kullanılmıştır. Üzerinde çalışılan tüm algoritmalar ile k-katlamalı çapraz doğrulama (k=10) kullanılarak çalışılmıştır. Bu yöntem ile çalışma kapsamında eksik öğrenme ve aşırı öğrenme sorunlarının önüne geçilmiştir. Altı farklı algoritmaya ait sonuçlar Çizelge-4'te sunulmuştur.

Çizelge-4 ve Şekil-9 incelendiğinde tutturma, f-değeri ve doğruluk değerleri açısından en iyi sonucu Rassal Orman algoritmasının sağladığı görülmektedir. Bu sınıflandırma algoritması Karar Ağacı algoritması temeline dayanmaktadır. Bu durum Çizelge-4 üzerindeki verilerden de gözlenebilmektedir. Rassal Orman Sınıflandırıcısı sonrasında en iyi sonucu Karar Ağacı Sınıflandırıcısı vermiştir. Bu noktada Rassal Orman Sınıflandırıcısı, Karar Ağacı Sınıflandırıcısı'nın en büyük problemi olan aşırı öğrenmenin önüne geçmek için hem veri kümesi hem de öznelikler üzerinde rassal olarak farklı alt setler seçmekte ve bunları eğitmektedir. Fazla sayıda oluşan karar ağaçlarını bağımsız olarak çalıştırmakta ve en uygununu seçmektedir. Veri seti içerisinde bulunan 16 farklı öznelik birbirinden bağımsız nitelikte veriler içermesi ve sonuç olarak ikili bir kategorik veriye ulaşması açısından ilk başta karar ağacının yüksek performans elde edilebileceği öngörülmüştür. Ancak bu durum aşırı öğrenmenin önüne geçilemek ve alt setler üzerinde yeni karar ağaçlarının oluşturulması ve en iyi sonucu veren karar ağacının uygulanmasını sağlayan Rassal Orman algoritmasının daha iyi sonuç üretmesini sağlamıştır.

İstihdam işlemleri gibi kritik öneme sahip durumlarda son yıllarda kullanılan karar destek sistemlerine paralel olarak belirli bir küme içerisinde en uygun adayın seçilmesine yönelik yapay öğrenme destekli bu çalışma ile maddi hataların önüne geçilmesi ve belirlenen kısıtlara en uygun adayın belirlenmesi sağlanmıştır. Çalışmanın sonucunda kullanılan algoritmalar ile 0.81 doğruluk oranı gibi yüksek düzeyde olumlu sonuç elde edilmiştir. Gelecek araştırmalarda farklı öznelikler içeren veri setlerinin kullanımı ve daha fazla sınıflandırıcı algoritmaların kullanımını içeren çalışmalar planlanmaktadır.

Kaynakça

- [1] Azar, A. "Türkiye'deki Öğretmen Eğitimi Üzerine Bir Söylem: Nitelik mi, Nicelik mi?", Yükseköğretim ve Bilim Dergisi, sy 1, Art. sy 1, Nis. 2011.
- [2] Chubak J. ve ark., "Tradeoffs between accuracy measures for electronic health care data algorithms", Journal of Clinical Epidemiology, c. 65, sy 3, ss. 343-349.e2, Mar. 2012, doi: 10.1016/j.jclinepi.2011.09.002.
- [3] Özgöbek Ö. ve ark. "A Survey on Challenges and Methods in News Recommendation", Proceedings of the 10th International Conference on Web Information Systems and Technologies, Barcelona, Spain, 2014, ss. 278-285. doi: 10.5220/0004844202780285.
- [4] Sayar A. ve Turdaliev N., "Makine Öğrenmesi ile Adaptif Otel Öneri Sistemi", 12th Turkish National Software Engineering Symposium, Istanbul, Türkiye, s. 13.
- [5] Dündar A. ve Kakişim A., "Kıyafet Öneri Sistemi için Giyim Metaverilerine dayalı Temsil Öğrenimi", European Journal of Science and Technology, Ara. 2021, doi: 10.31590/ejosat.1008736.
- [6] Bal M., ve ark., "İnsan Kaynakları Yönetiminde Etkin Bir İşe Alım Süreci İçin Yapay Zeka Yöntemlerinin Kullanımı", 3rd International Congress of Eurasian Social Sciences, 2019.
- [7] Gültekin B. ve ark., "İşe Alım Süreci Aday Ön Tesbitinde Bulanık Mantık Tabanlı SQL Sorgulama Yönteminin İncelenmesi", Düzce Üniversitesi Bilim ve Teknoloji Dergisi, c. 3, sy 1, ss. 198-209, 2015.
- [8] Büyükoztürk Ş., "Faktör Analizi: Temel Kavramlar ve Ölçek Geliştirmede Kullanımı", Kuram ve Uygulamada Eğitim Yönetimi, c. 32, sy 32, Art. sy 32, Mar. 2002.
- [9] Rodríguez P. ve ark., "Beyond One-Hot Encoding: Lower Dimensional Target Embedding", Image and Vision Computing, c. 75, ss. 21-31, Tem. 2018, doi: 10.1016/j.imavis.2018.04.004.
- [10] Swinburne R., "Bayes' Theorem", Revue Philosophique de la France Et de l'Étranger, c. 194, sy 2, ss. 250-251, 2004, doi: 10.1093/mind/113.451.591.
- [11] Shi T. ve Horvath S., "Unsupervised Learning With Random Forest Predictors", Journal of Computational and Graphical Statistics, c. 15, sy 1, ss. 118-138, Mar. 2006, doi: 10.1198/106186006X94072.
- [12] T. S. Wiens ve diğ., "Three way k-fold cross-validation of resource selection functions", Ecological Modelling, c. 212, sy 3-4, ss. 244-255, Nis. 2008, doi: 10.1016/j.ecolmodel.2007.10.005.
- [13] T.T. Wong ve N.Y. Yang, "Dependency analysis of accuracy estimates in k-fold cross validation", IEEE Transactions on Knowledge and Data Engineering, c. 29, sy 11, ss. 2417-2427, 2017.
- [14] S. Arlot ve A. Celisse, "A survey of cross-validation procedures for model selection", Statistics Surveys, c. 4, sy none, ss. 40-79, Oca. 2010, doi: 10.1214/09-SS054.