

Veri Madenciliği Teknikleri Kullanılarak Lise Öğrencilerinin Mesleki Kişilik Profilleri ve Kişilik Özelliklerine Dayalı Kariyer Yönlendirme: Makine Öğrenmesi

Araştırma Makalesi/Research Article

 Murat URFALIOĞLU^{1*},  Nuray AŞANTUĞRUL²

¹Yönetim Bilişim Sistemleri Bölümü, Bolu Abant İzzet Baysal Üniversitesi, Bolu, Türkiye

²Eğitim Bilimleri Bölümü, Amasya Üniversitesi, Amasya, Türkiye

murat.urfalioglu@ibu.edu.tr, nuray.asantugrul@amasya.edu.tr

(Geliş/Received:14.06.2025; Kabul/Accepted:12.11.2025)

DOI: 10.17671/gazibtd.1719617

Özet— Bu çalışmanın amacı, lise öğrencilerini mesleki kişilik profilleri, beş faktör kişilik özellikleri ve demografik özelliklerine göre uygun fakültelere yönlendirecek makine öğrenmesi tabanlı bir sınıflandırma modeli geliştirmektir. Modelde genel not ortalaması 2.5 ve üzeri olan üniversite öğrencilerinden elde edilen veriler kullanılmıştır. Üniversite öğrencileri, lisans programlarını seçmiş ve bu alanda belirli bir başarı düzeyine ulaşmış olmaları nedeniyle, lise öğrencilerinin potansiyel mesleki yönelimlerini tahmin edebilecek bir grup olarak değerlendirilmiştir. Böylece geliştirilen modelin, lise düzeyinde uygulanabilecek öngörüsül bir yönlendirme aracı olarak kullanılabilmesi amaçlanmıştır. Araştırmanın çalışma grubunu 2024-2025 eğitim-öğretim yılında farklı fakültelerde öğrenim görmekte olan ve uygun örnekleme yöntemiyle belirlenen 877 üniversite öğrencisi (502’si kadın, 375’i erkek) oluşturmaktadır. Araştırmada veriler Kişisel Bilgi Formu, Beş Faktör Kişilik Ölçeği ve Mesleki Kişilik Tipleri Envanteri ile toplanmıştır. Elde edilen veri setinde toplam 51 öznitelik bulunmaktadır. WEKA 3.9.5 yazılımı kullanılarak Random Forest, SMO (SVM), Naive Bayes, J48 ve IBk (k-NN) algoritmaları uygulanmış ve performans ölçütleri karşılaştırılmıştır. Bulgular, Random Forest algoritması %87.6 doğruluk ve 0.867 F1 skor değeri ile en başarılı model olduğunu göstermiştir. Ayrıca öznitelik önem analizi sonuçları, fakülte yerleşiminde aile özellikleri, akademik geçmiş ve mesleki kişilik tiplerinin belirleyici değişkenler olduğunu ortaya koymuştur. Araştırma sonuçları, yapay zekâ destekli öğrenci yönlendirme sistemlerinin geliştirilmesine ve veri temelli kariyer danışmanlığı uygulamalarına katkı sağlamaktadır.

Anahtar Kelimeler— makine öğrenmesi, fakülte tahmini, kariyer yönlendirme, veri madenciliği

Career Guidance Based on Vocational Personality Profiles and Personality Traits of High School Students Using Data Mining Techniques: Machine Learning

Abstract— The aim of this study is to develop a machine learning-based classification model to guide high school students toward appropriate faculties based on their vocational personality profiles, Big Five personality traits, and demographic characteristics. The model was trained using data obtained from university students with a grade point average of 2.5 or higher. Since these students have already selected a bachelor’s program and achieved a certain level of academic success, they were considered a suitable group for predicting potential vocational orientations at the high school level. The study sample consisted of 877 university students (502 female, 375 male) enrolled in different faculties during the 2024–2025 academic year and selected through convenience sampling. Data were collected using a Personal Information Form, the Big Five Personality Scale, and the Vocational Personality Types Inventory, resulting in 51 attributes. Using WEKA 3.9.5, Random Forest, SMO (SVM), Naive Bayes, J48, and IBk (k-NN) algorithms were applied and compared. The findings showed that Random Forest achieved the best performance, with an accuracy of 87.6% and an F1 score of 0.867. Feature importance analysis indicated that family characteristics, academic background, and vocational personality types were the most influential factors. These results support the development of AI-based student guidance systems and data-driven career counseling applications.

Keywords— machine learning, faculty prediction, career guidance, data mining

1. GİRİŞ (INTRODUCTION)

Kariyer seçimi, bireyin yaşamını şekillendiren kritik kararlardan biridir. Bu nedenle bireyin kariyer seçimini yaparken kendi kişisel özellikleri ile seçeceği mesleğin özellikleri arasında uyum olmasına dikkat etmesi önem taşımaktadır [1]. Bireylerin doğru meslek seçimleri yapabilmesi, ilgileri, kişilik özellikleri, akademik başarıları gibi özelliklerinin birlikte değerlendirilmesini gerektirmektedir. Bu noktada kişilik özellikleri, kariyer seçimini belirleyen temel faktörlerden biri olarak öne çıkmaktadır.

Kişilik kavramı eskiden beri sıklıkla çalışılan ve araştırmalara konu olan bir olgudur. Yapılan literatür taramalarında, kişilik kavramına ilişkin çok sayıda tanımın yer aldığı görülmektedir. Kişilik; bireyin kalıtsal özellikleri ile çevresel etkileşimler sonucunda şekillenen ve kolaylıkla değişmeyen bir özellikler bütünü olarak tanımlanabilir. [2]. Köroğlu ve Bayraktar'a [3] göre kişilik, bireyin ayırt edici özelliklerinin ve kendine göre bir farklılığının olması durumudur. Bireyin kişiliğinin oluşmasında kalıtım ile çevre ve toplumsal faktörler etkili olmaktadır [4].

Kişiliğin değerlendirilmesinde kullanılan kuramlardan biri beş faktör kişilik modelidir. Bu model, kişilikle ilgili çalışmalar yapan araştırmacılar tarafından sıklıkla kullanılmaktadır ve bunun nedeni olarak modelin şu beş özelliği rol oynamaktadır. İlk olarak modelin ampirik ve boylamsal çalışmalara dayanması, bazı biyolojik temellere dayanması, çeşitli kültürlerde geçerliliğinin sağlanması, ölçülen özelliklerin sürekliliğini koruması ve son olarak psikometrik açıdan değerlendirilmesinin kolay olmasıdır [5]. Beş Faktör Kişilik Modeli'nde, bireylerin hem kendilerini hem de başkalarını tanımlarken kullandıkları sıfatlar kişiliğin değerlendirilmesinde temel alınmaktadır. Bu yönüyle modelin, özellik yaklaşımını temel aldığı ifade edilebilir. Bu modelde kişilik; deneyime açıklık, dışadönüklük, öz-denetim, yumuşak başlılık ve duygusal denge (nörotizm) alt boyutlarından meydana gelmektedir [6]. Son yıllarda bu model, bireylerin kariyer tercihleriyle ilişkisini açıklamak amacıyla pek çok çalışmada kullanılmıştır [7, 8, 9, 10].

Kariyer seçiminde kişilik özelliklerini merkeze alan kuramlardan biri de Holland'ın Tipoloji Kuramı'dır. Bu kuramın kullanışlı, sade ve net olması, kişilik özellikleri ile iş çevresi arasındaki ilişkiyi belirgin biçimde yapılandırılması sayesinde günümüzde de geçerliliğini korumakta ve kariyer danışmanlığı uygulamalarında sıkça kullanılmaktadır. [11, 12]. Holland, bireylerin mesleki kişilik özelliklerini ve bu özelliklerin uygun olduğu mesleki ortamları açıklamak amacıyla altı temel kişilik tipi ve buna karşılık gelen mesleki çevre modelini ortaya koymuştur. Bu tipler; gerçekçi (realistic), araştırmacı (investigative), sanatçı (artistic), sosyal (social), girişimci (enterprising) ve geleneksel (conventional) olarak sıralanır. Bu model, baş harflerinden oluşan "RIASEC" kısaltmasıyla da bilinmektedir. Her bir kişilik tipi; bireyin genel eğilimlerini, davranış kalıplarını, iş yapma tarzını ve

problem çözme yöntemlerini yansıtır. Holland'a göre bireyler genellikle bu altı kişilik tipinden birine daha yakındır ve kendilerine en uygun olan, kişilik özelliklerini yansıtan mesleki çevrelerde yer alma eğilimindedirler [13].

Holland'ın Tipoloji Kuramına göre insanlar altı kişilik tipi ve mesleki çevreye dair özellikler genel olarak şu şekilde tanımlanmıştır. Gerçekçi tip makineler, aletler ve nesnelerle çalışmaktan hoşlanmaktadır. Mekanik becerileri yüksektir, tamir etme ve araç kullanma gibi işlerle meşgul olmayı severler [14, 15]. Araştırmacı tip olayları nedenlerini araştırmak, kontrol etmek veya ortaya çıkartmak amacıyla gözleme dayalı, yaratıcı ve sistematik araştırma yapmayı tercih ederler [16]. Sanatçı tip bireyler özgün olmaktan hoşlanırlar, duygu ve düşüncelerini yaratıcı yöntemlerle (resim, müzik, edebiyat gibi) ifade etmeyi severler [14]. Bu tipteki bireyleri tanımlamada yaratıcı, etkileyici, sezgisel, dürtüsel ve bağımsız gibi sıfatlar uygundur [17]. Sosyal tip bireyler dışadönük ve empati becerisi güçlü yapıdadırlar. Bu tip bireyler insanlarla iş birliği yapılabilecek iletişime geçilebilecek ve başkalarına yardım edilebilecek iş ortamlarında çalışmayı tercih ederler [15]. Girişimci tip bireyler maddi kazanç ve sosyal statüye önem vermektedirler [14]. Bu tip bireylerin temel özellikleri dışadönük, girişken, enerjik, ikna yeteneği ve kendine güveni yüksek olmaktadır. Topluma liderlik etmekten keyif alırlar [18]. Son olarak geleneksel tip bireyler sistematik, düzenli ve kesin sonuçları olan işlerde çalışmayı tercih ederler. Esnek, plan içermeyen, belirsiz ve yapılandırılmamış ortamlarda çalışmayı tercih etmezler [16].

Teknolojik dönüşümle birlikte kariyer danışmanlığı süreçlerinde veri madenciliği ve makine öğrenmesi gibi yöntemlerin kullanımı hızla artmıştır. Makine öğrenmesi, bir makineye nasıl öğreneceğini öğreten yapay zekanın belirli bir dalı olup veriden otomatik olarak öğrenme ve tahmin yapma yeteneğiyle günümüzün en önemli teknolojik alanlarından biridir. Makine öğrenmesi; eğitim başta olmak üzere sağlık, üretim, finans, pazarlama, uzaktan algılama ve tedarik zinciri gibi birçok alanda kullanılmaktadır [19, 20, 21].

Makine öğrenmesi ve veri madenciliği tekniklerinin eğitim alanında kullanılmasıyla birlikte, öğrencilere daha kişiselleştirilmiş kariyer rehberliği sunmak mümkün hale gelmiştir. Mevcut rehberlik hizmetleri veya kariyer seçimi doğrultusunda bireylerin yönlendirilmesi çoğunlukla geleneksel yöntemlerle yürütülmekte olup, öğrencilerin bireysel mesleki özelliklerini sistematik biçimde analiz eden veri odaklı modellere sınırlı biçimde yer verilmektedir. İlgili literatür incelendiğinde makine öğrenmesi ve veri madenciliği yönteminin kariyer danışmanlığı alanında henüz son yıllarda çalışmaya başlandığı görülmektedir. Örneğin Song vd. (2024) [22] kariyer seçiminin tahmininde makine öğrenmesini incelemişlerdir. Hem çalışan hem işsiz katılımcılardan oluşan büyük bir örneklem ile yürütülen bu çalışmada geleneksel profil yöntemine kıyasla, makine öğrenimi destekli yöntemin her iki kariyer seçeneği türünü de tahmin etmede daha yüksek genel doğruluk sağladığı sonucuna

ulaşmıştır. Bahalkar vd. (2025) [23] öğrencilerin akademik performanslarına, ders dışı etkinliklerine ve kişisel isteklerine dayanarak potansiyel kariyer yollarını tahmin etmek için tasarlanmış yapay zeka tabanlı bir öngörü modeli geliştirmişlerdir. Model, öğrenciler için en uygun ders seçimlerini anlama ve tahmin etme yeteneğini gösteren yüksek bir doğruluk oranına ulaşmıştır. Benzer olarak Shahane vd. (2024) [24], makine öğrenimini kullanarak öğrencilerin kariyer ilgi alanları tahmini üzerine bir çalışma yürütmüşlerdir. Türkiye’de yapılan çalışmalar incelendiğinde ise benzer yaklaşımlar oldukça sınırlı olduğu, veri madenciliği ve makine öğrenmesinin kariyer danışmanlığına entegrasyonunun henüz yeni olduğu görülmektedir [25, 26, 27]. Bu çalışmalar önemli katkılar sunmakla birlikte, çoğu yalnızca kişilik tiplerine ya da akademik başarıya odaklandığı görülmüştür. Türkiye bağlamında lise öğrencileri üzerinde yürütülmüş, hem Beş Faktör Kişilik Modeli hem Holland Tipolojisi temelli makine öğrenmesi sınıflandırmalarına rastlanmamıştır. Ülkemizde kariyer danışmanlığı alanında veri madenciliği yöntemi kullanarak bireylerin doğru kariyer kararı vermelerine yardım edebilmek önemli bir ihtiyaçtır. Bu doğrultuda bu araştırmanın amacı, öğrencilerin mesleki kişilik profilleri, beş faktör kişilik özellikleri ve demografik değişkenlerine dayalı olarak fakülte yerleşimlerini tahmin edebilen bir makine öğrenmesi modeli geliştirmektir. Bu modelin geliştirilmesiyle, bireylere kişisel özelliklerine uygun kariyer alternatiflerinin veri temelli biçimde önerilmesi hedeflenmektedir.

Bu temel amaç çerçevesinde aşağıdaki alt araştırma sorularına yanıt aranmaktadır:

- Hangi sınıflandırma algoritması, öğrencilerin fakülte yerleşimlerini en yüksek doğrulukla tahmin etmektedir?
- Hangi öznitelikler fakülte tahmininde en yüksek etkiye sahiptir?
- Sınıflandırma algoritmaları sınıf etiketi (yerleşilen fakülte) temelinde en başarılı ve en başarısız hangi sınıflarda performans göstermektedir?

2. YÖNTEM (METHOD)

Bu çalışmada, lise düzeyindeki öğrencilerin kişilik, ilgi ve başarı özelliklerine dayanarak hangi fakültede başarılı olabilecekleri tahmininde bulunan bir sınıflandırma modeli geliştirmek amacıyla makine öğrenmesi algoritmaları ve veri madenciliği yöntemlerinden yararlanılmıştır. Veri madenciliği, büyük veri kümelerinde anlamlı, beklenmeyen veya değerli örüntülerin ve ilişkilerin keşfedilmesidir [28]. Veri madenciliği; makine öğrenmesi, desen tanıma, istatistik, veri tabanı sistemleri ve görselleştirme alanlardan yararlanarak, geniş veri kümeleri içinde anlamlı ve önceden bilinmeyen ilişkileri ortaya çıkarmayı amaçlayan disiplinler arası bir alandır [29]. Bu teknik, etiketlenmiş örnek verilerden öğrenerek, sonuçları

bilinmeyen yeni verilerin sınıflandırılması veya tahmin edilmesi amacıyla kullanılabilir [30].

2.1. Çalışma Grubu (Study Group)

Araştırmanın çalışma grubu uygun örnekleme yöntemiyle belirlenen katılımcılardan oluşmaktadır. Bu bağlamda Tokat Gaziosmanpaşa Üniversitesi ve Amasya Üniversitesi’nin farklı fakültelerinde 2024–2025 eğitim öğretim yılında öğrenim görmekte olan ve genel not ortalaması (GNO) 4’lük sisteme göre 2.5 ve üzeri olan toplam 877 lisans öğrencisi oluşturmaktadır. Uygun örnekleme yöntemi, olasılığa dayalı olmayan ve araştırmacının kolay erişilebilirlik, maliyet, zaman gibi etkenleri göz önünde bulundurarak çalışmaya dahil edilecek gruba belirlediği bir örnekleme yöntemidir [31]. Araştırma kapsamında veriler üniversite öğrencilerinden gönüllülük esasına göre toplanmıştır. Araştırmanın örneklemini, Tokat Gaziosmanpaşa Üniversitesi ve Amasya üniversitelerindeki 15 farklı fakülteden ve her bir fakülteden yaklaşık 60 öğrenciden oluşmaktadır.

Bu araştırmanın temel amacı, lise düzeyindeki öğrencilerin kişilik, ilgi ve başarı özelliklerine dayanarak hangi fakültede başarılı olabilecekleri tahmininde bulunan bir sınıflandırma modeli geliştirmektir. Ancak, sınıflandırma modelinin güvenilir ve geçerli bir şekilde eğitilebilmesi için, mevcut durumda herhangi bir fakülteye yerleşmiş ve başarılı olmuş öğrencilerin verilerine ihtiyaç duyulmaktadır. Bu nedenle, çalışma kapsamında veri toplama evreni olarak farklı fakültelerde öğrenim görmekte olan ve genel not ortalaması (GNO) 2.5 ve üzeri olan üniversite öğrencileri belirlenmiştir. Makine öğrenmesi temelli sınıflandırma algoritmalarının eğitilebilmesi için, gözlemlere ait sınıf etiketlerinin (ground truth) önceden belirlenmiş olması gerekmektedir [32]. Lise öğrencilerinin gelecekte hangi fakülteye yerleşecekleri bilinemezken, üniversite öğrencileri için bu etiket (örneğin Diş Fakültesi) hâlihazırda tanımlıdır. Bu durum, algoritmanın doğruluğunun değerlendirilebilmesi açısından belirleyici bir unsurdur.

2.2. Veri Toplama Araçları (Data Collection Tools)

Veri toplama sürecinde, [33] tarafından geliştirilen “Beş Faktör Kişilik Ölçeği (BFKÖ), [34] tarafından geliştirilen RIASEC Mesleki Kişilik Envanteri ile araştırmacılar tarafından geliştirilen öğrencilerin demografik özellikleri (cinsiyet, yaş, lise türü, aile gelir düzeyi, anne-baba eğitim düzeyi) ile akademik performans verilerinin (ortaokul, lise notu, üniversite GNO) yer aldığı Kişisel Bilgi Formu kullanılmıştır. Özniteliklerin dağılımı Tablo 1’de verilmiştir.

Tablo 1. Öznitelik sayısı dağılımı
(Table 1. Attribute count distribution)

Öznitelik	Öznitelik sayısı
BFKÖ	30
RIASEC Mesleki Kişilik Envanteri	10

Akademik performans verileri	3
Öğrenci demografik özellikleri	7
Sınıf etiketi (Yerleşilen Fakülte)	1
Toplam Öznitelik	51

2.2.1. Beş Faktör Kişilik Ölçeği (BFKÖ) (Five-Factor Personality Inventory (FFPI))

Beş Faktör Kişilik Ölçeği, beş temel kişilik özelliğini ölçmek üzere [35] tarafından geliştirilmiştir. Ölçeğin Türk kültürüne uyarlaması [33] tarafından yapılmıştır. Ölçek formu 10 madde ve beş alt boyuttan oluşmaktadır. Bu alt boyutlar; dışadönüklük, yumuşak başlılık, öz denetimlilik, nörotiklik ve deneyime açıklıktır. Geçerlik analizi kapsamında yapılan doğrulayıcı faktör analizi ile ölçeğin uyum indeksleri incelenmiş ve minimum ki-kare değerinin ($\chi^2=46.139$, $p=0.00$) anlamlı olduğu görülmüştür. Ayrıca uyum indeksi değerleri GFI=.96, AGFI=.91, NFI=.97, CFI=.98, RMSEA=.062 ve SRMR=.035 olarak elde edilmiştir. Uyum indekslerine yönelik ölçütler dikkate alındığında belirtilen indeksler için mükemmel uyum olduğu görülmüştür [36]. Güvenirlik analizi sonucunda ise ölçeğin beş faktörü için iç tutarlılık ve kompozit güvenilirlik değerleri sırasıyla dışadönüklük için .88 ve .83; yumuşak başlılık için .81 ve .73; öz denetimlilik için .90 ve .85, nörotiklik için .85 ve .79 ve deneyime açıklık için .84 ve .78 güvenilirlik değerine sahip olduğu bulunmuştur. İç tutarlılık ve kompozit güvenilirlik değerlerinin .70'den yüksek olması ise ölçeğin güvenilirlik değerlerinin yüksek olduğunu göstermektedir. Son olarak ölçekte “Kendimi içine kapanık biri olarak görüyorum”, “Kendimi kolay sınırlanan biri olarak görüyorum” gibi maddeler bulunmaktadır.

2.2.2. Mesleki Kişilik Tipleri Envanteri (Vocational Personality Types Inventory)

Mesleki Kişilik Tipleri Envanteri, [34] tarafından Holland'ın kuramına dayalı olarak geliştirilmiştir. Envanter, 30 madde ve 6 alt boyuttan oluşmaktadır. Alt boyutlar araştırmacı, gerçekçi, sosyal, girişimci, geleneksel ve sanatçı kişilik tipleri olarak adlandırılmıştır. Geçerlik analizi kapsamında gerçekleştirilen açımlayıcı faktör analizi sonucunda madde faktör yüklerinin .93 ile .45 arasında değer aldığı görülmüştür. Doğrulayıcı faktör analizi sonucunda elde edilen bulgular incelendiğinde χ^2/sd oranı 1.730 ($\chi^2/sd=662.670/383$) olduğu görülmektedir. [37], χ^2/sd oranının 2'den küçük değerler alması iyi uyumu, 2 ile 3 arasındaki değerler alması kabul edilebilir uyumu gösterdiğini belirtmişlerdir. Bu bağlamda, elde edilen bu χ^2/sd oranının iyi uyumu gösterdiği söylenebilir. Güvenirlik analizi kapsamında Cronbach Alfa değerleri girişimci alt boyutu için .85, gerçekçi alt boyutu için .72, sosyal alt boyutu için .81, araştırmacı alt boyutu için .81, geleneksel alt boyutu için .78 ve sanatçı alt boyutu için .77 olarak elde edilmiştir. Ayrıca yapılan test tekrar test uygulama sonucunda envanterin alt boyutlarına ilişkin korelasyon değerlerinin .62 ile .80 arasında değiştiği görülmektedir. Elde edilen bu değerler doğrultusunda ölçeğin mesleki ilgi alanlarını güvenilir ve geçerli bir biçimde sınıflandırmak amacıyla geliştirildiği söylenebilir.

Ölçekte “Laboratuvar ortamında deneyler yapmak isterim”, “İnsanların sorunlarını çözmelerine yardım etmek isterim” gibi maddeler bulunmaktadır.

2.2.3. Kişisel Bilgi Formu (Personal Information Form)

Öğrencilerin demografik özellikleri (yaş, cinsiyet, aile gelir düzeyi, lise türü, anne-baba eğitim düzeyi) ile akademik performans verileri (ortaokul, lise notu, üniversite GNO) gibi bilgilerin alındığı araştırmacılar tarafından oluşturulan formdur.

2.3. Verilerin Toplanması ve Analizi (Data Collection and Analysis)

Araştırmanın veri seti Tablo 1’de verilmiş olup 877 öğrenciden 51 öznitelikte veri toplayarak veri seti oluşturulmuştur. Veriler, her biri yaklaşık 60 öğrenciden oluşan 15 farklı fakültede öğrenim gören öğrencilerden toplanmıştır. Veri setindeki tüm veriler eksik değer kontrolünden geçirilmiş ve sayısal formatlara uygun biçimde kodlanarak WEKA programında sayısal ya da nominal dönüşümleri yapılmıştır. Model belirlemede girdiler olan RIASEC, BFKÖ, akademik performans verileri sayısal; demografik özellikler değişkenleri nominal, sınıf değişkeni ise “yerleşilen_fakulte” nominal olarak belirlenmiştir.

Bu çalışmada, veri setinin analizinde, veri madenciliği ve makine öğrenmesi alanlarında yaygın olarak kullanılan açık kaynaklı bir yazılım olan WEKA (Waikato Environment for Knowledge Analysis) programından yararlanılmıştır. WEKA, Java tabanlı bir platformdur. Kullanıcıya birçok sayıda makine öğrenmesi algoritması ile veri ön işleme, sınıflandırma, kümeleme, regresyon ve görselleştirme gibi işlemlerin yapılmasını sağlar. Ayrıca, model doğrulama ve performans değerlendirme araçları (örneğin doğruluk, kappa istatistiği, hata oranları) ile algoritmaların etkinliğini karşılaştırmayı sağlamaktadır. Bu özellikleriyle WEKA, akademik araştırmalarda sıkça tercih edilen bir veri madenciliği ve makine öğrenmesi aracıdır [38]. WEKA programında veri seti 5 farklı sınıflandırma algoritmaları kullanılarak model performansları karşılaştırılmıştır. Kullanılan sınıflandırma algoritmaları Tablo 2’de verilmiştir.

Tablo 2. Kullanılan sınıflandırma algoritmaları ve açıklamaları

(Table 2. Used classification algorithms and their explanations)

No	Algoritma Adı	WEKA Konumu	Açıklama
1	Decision Tree (J48)	trees > J48	C4.5 algoritmasının Weka'daki uygulaması olup karar ağacı algoritmasıdır. Veriyi karar kuralları ile dallandırarak sınıflandırır [39].

2	Random Forest	trees > RandomForest	Karar ağaçlarının bir topluluğunu oluşturarak sınıflandırma ve regresyon problemlerinde yüksek doğruluk sağlayan bir makine öğrenmesi yöntemidir [40].
3	Naive Bayes	bayes > NaiveBayes	Bayes teoremine dayanan ve özniteliklerin bağımsız olduğu varsayımını kullanan Bayes ağı sınıflandırıcısıdır [41].
4	IBk (Instance-Based k-nearest neighbors, yani k-NN)	lazy > IBk	IBk, k-en yakın komşu sınıflandırıcısının Weka uygulamasıdır. Tüm eğitim verilerini depolayan ve tahmin sırasında kullanan, genellikle varsayılan metrik olarak Öklid mesafesini kullanan örnek tabanlı bir öğrenicidir [42].
5	Support Vector Machine (SVM)	functions > SMO	Sınıfları en iyi ayıran karar sınırını (hyperplane) bulmayı amaçlar. [43]. Küçük örneklemelerde yüksek doğruluk sağlar.

Tablo 2’de verilen algoritmaların oluşturduğu modeller, doğruluk (accuracy), kappa katsayısı (kappa statistic), F1 skoru (F1-score), kesinlik (precision) ve duyarlılık (recall) ölçütlerinin yanı sıra, ortalama mutlak hata (Mean Absolute Error) ve kök ortalama kare hata (Root Mean Square Error) hata metrikleri temel alınarak değerlendirilmiştir. Ayrıca, her bir sınıflandırma algoritmasına ait karmaşıklık matrisi (confusion matrix) sonuçları da analiz edilmiştir. Bu kapsamda, modellerin performansı ayrıntılı biçimde incelenmiştir. Tüm algoritmalarda örnekleme yanlılığını en aza indirmek amacıyla 10 katlı çapraz doğrulama (10-fold cross-validation) yöntemi kullanılmıştır. Kullanılan

algoritmalara ait parametre ayarları ve açıklamaları aşağıda özetlenmiştir.

Decision Tree (J48), C4.5 karar ağacı algoritmasının WEKA’daki uygulamasıdır.

Kullanılan parametreler: *weka.classifiers.trees.J48 -C 0.25 -M 2*

- C 0.25: Budama işlemi için güven düzeyidir. Değer küçüldükçe daha fazla budama gerektirir.
- M 2: Bir yaprak düğümünde bulunabilecek minimum örnek sayısını ifade etmektedir.

Random Forest, birden fazla karar ağacının (Decision Tree) bir araya gelmesiyle oluşturulan topluluk yöntemidir.

Kullanılan parametreler: *weka.classifiers.trees.RandomForest -P 100 -attribute-importance -I 100 -num-slots 1 -K 0 -M 1.0 -V 0.001 -S 1*

- P 100: Her bir yinelemede (iteration) örnekleme oranı %100 olarak belirlenmiştir.
- attribute-importance: Modelin her bir öznitelik için bilgi kazancı esaslı önem düzeylerini hesaplamasını sağlar.
- I 100: 100 adet karar ağacından (tree) oluşan bir topluluk (ensemble) modeli oluşturur.
- num-slots 1: Ensemble oluşturulması sırasında kullanılacak yürütme yuvalarının (iş parçacıklarının) sayısıdır.
- K 0: Her düğümde tüm değişkenler değerlendirilmiştir (özellik altkümesi seçimi yapılmamıştır).
- M 1.0: Yaprak düğümlerinde bulunması gereken minimum örnek sayısı 1 olarak ayarlanmıştır.
- V 0.001: Doğruluk iyileştirme eşiği 0.001 olarak belirlenmiştir; modelin durma kriteri için alt sınırdır.
- S 1: Rastgelelik sabiti (random seed) 1 olarak belirlenmiş ve böylece tekrarlanabilir sonuçlar elde edilmiştir.

NaiveBayes, örneklerin belirli bir sınıfa ait olma olasılığını hesaplamak için koşullu olasılık yaklaşımını kullanır.

Kullanılan parametreler: *weka.classifiers.bayes.NaiveBayes*

- BatchSize: İşlenmesi tercih edilen örnek (instance) sayısını belirtir. 100 olarak varsayılan değerdir.

- NumDecimalPlaces: Modeldeki sayısal çıktılar için kullanılacak ondalık basamak sayısını belirtir. 2 olarak varsayılan değerdir.

IBk (k-Nearest Neighbors), örnek tabanlı sınıflandırma yöntemlerinden biridir. Algoritma, sınıflandırma yapılacak örneğin en yakın “k” komşusunu belirleyerek, bu komşuların sınıf etiketlerine göre tahminde bulunur.

Kullanılan Parametreler: *IBk -K 1 -W 0 -A "weka.core.neighboursearch.LinearNNSearch -A "\"weka.core.EuclideanDistance -R first-last\""*

- K 1: Komşu sayısıdır. k=1 olarak belirlenmiştir. Yani sınıflandırma, en yakın tek komşuya göre yapılır.
- W 0: Ağırlıklandırma seçeneğidir. 0 değeri, eşit ağırlıklandırma anlamına gelir (her komşu eşit öneme sahiptir).
- A
"weka.core.neighboursearch.LinearNNSearch": Komşuluk arama yöntemi, Lineer arama kullanılır; yani algoritma her örneği tek tek karşılaştırarak en yakın komşuları bulur.
- A "weka.core.EuclideanDistance -R first-last": Mesafe ölçütü Öklidyen uzaklık (Euclidean Distance) kullanılmıştır. -R first-last ifadesi, tüm özniteliklerin (ilk sütundan son sütuna kadar) mesafe hesabına dahil edildiğini belirtir.

Support Vector Machine (SVM), veri kümelerindeki sınıfları maksimum marjini sağlayan bir hiper düzlem (hyperplane) aracılığıyla ayırmaktır.

Kullanılan parametreler: *SMO -C 1.0 -L 0.001 -P 1.0E-12 -N 0 -V -1 -W 1 -K "weka.classifiers.functions.supportVector.PolyKernel -E 1.0 -C 250007" -calibrator "weka.classifiers.functions.Logistic -R 1.0E-8 -M -1 -num-decimal-places 4"*

- C 1.0: Karmaşıklık parametresidir (Complexity parameter). Modelin karmaşıklığını ve hata toleransını kontrol eder.
- L 0.001: Tolerans değeridir (tolerance parameter).
- P 1.0E-12: Yuvarlama hatası için Epsilon değeridir. Bu değerin değiştirilmemesi önerilir. Varsayılan olarak gelmektedir.
- N 0: Normalize edilip edilmeyeceğidir. 0 değeri, verilerin normalize edilmediğini belirtir.
- V -1: Kalibrasyon modelleri için eğitim verisi üretmek amacıyla kullanılan çapraz doğrulama katlama sayısıdır. Değerin -1 olması eğitim verisi kullanıldığı anlamına gelir.

- W 1: çapraz doğrulama (cross-validation) sürecinde kullanılan rastgele sayı tohumu (random seed) değerini belirtir.
- K: Kullanılacak çekirdek fonksiyonunu (kernel) belirtir. Polinomsal çekirdek (Polynomial Kernel) kullanılmıştır.
- Calibrator: Olasılık kalibrasyonudur. Çıktı olasılıklarını kalibre etmek için lojistik regresyon kullanılmıştır.

Karmaşıklık matrisi, çok etiketli sınıflandırıcıların davranışlarına dair özlü ve sistematik bir bakış açısı sunarak, bu sınıflandırma görevlerinde performans değerlendirmesini daha ayrıntılı ve açıklayıcı hâle getirmektedir [44]. Bu değerlendirme sınıf bazında modelin güçlü ve zayıf yönlerini ortaya koymak amacıyla yapılmıştır.

Ayrıca fakülte yerleştirmede en etkili öznitelikleri belirlemek için başarı performansı en yüksek çıkan Random Forest algoritması modelinin WEKA'da öznitelik önemi analizi (Feature Importance Analysis) yapılarak öznitelik önem puanları çıkarılmıştır.

Veri toplama sürecinde, 16 Mayıs 2025 tarihli ve E-30640013-108.01-261866 sayılı kararla Amasya Üniversitesi Sosyal Bilimler Etik Kurulu'ndan gerekli onay alınmıştır. Tüm katılımcılardan gönüllü onam formu ile izin sağlanmıştır. Elde edilen veriler doğrultusunda yalnızca başarılı öğrencilere (GNO ≥ 2.5) odaklanılmış, böylece analizde başarı eşliğini aşan öğrencilerin fakülte yerleşim özellikleri incelenmiştir.

3. BULGULAR (Findings)

Araştırma kapsamında elde edilen verilerden, katılımcıların demografik ve akademik özelliklerine ait istatistikler aşağıdaki tablolarda sunulmuştur.

Tablo 3. Cinsiyet dağılımı
(Table 3. Gender distribution)

Kategori	Frekans	Yüzde (%)
Kadın	502	57.2
Erkek	375	42.8
Toplam	877	100

Tablo 3'te 877 öğrencinin cinsiyete göre dağılımları verilmiştir. Tablo 3 incelendiğinde katılımcıların %57.2'si kadın (n = 502), %42.8'i ise erkek (n = 375) olduğu görülmektedir. Bu durum, veri setinde kadın öğrencilerin erkek öğrencilere kıyasla daha yüksek oranda temsil edildiğini göstermektedir.

Tablo 4. Anne öğrenim düzeyi dağılımı
(Table 4. Distribution of mother's education level)

Kategori	Frekans	Yüzde (%)
İlkokul	241	27.5
Ortaokul	210	24
Lise	232	26.5

Üniversite	194	22.1
Toplam	877	100

Tablo 4'te, 877 öğrencinin annelerine ait öğrenim düzeylerinin dağılımı verilmiştir. Araştırmaya katılan öğrencilerin annelerinin öğrenim düzeyine ilişkin veriler incelendiğinde, en yüksek oranın ilkökul mezunu annelere ait olduğu görülmektedir (n = 241, %27.5). Bunu sırasıyla lise mezunu (%26.5, n = 232), ortaokul mezunu (%24, n = 210) ve üniversite mezunu ebeveynler (%22.1, n = 194) izlemektedir. Veri setinde öğrencilerin annelerinin büyük çoğunluğu ortaöğretim ve altı düzeyde olduğu görülmektedir.

Tablo 5. Baba öğrenim düzeyi dağılımı
(Table 5. Distribution of father's education level)

Kategori	Frekans	Yüzde (%)
İlkokul	193	22
Ortaokul	252	28.7
Lise	214	24.4
Üniversite	218	24.9
Toplam	877	100

Tablo 5'te 877 öğrencinin babalarına ait öğrenim düzeylerinin dağılımı verilmiştir. Araştırmaya katılan öğrencilerin babaların öğrenim düzeyine ilişkin veriler incelendiğinde en yüksek oran ortaokul mezunu babalara aittir (n = 252, %28.7). Bunu sırasıyla üniversite mezunu (n = 218, %24.9), lise mezunu (n = 214, %24.4) ve ilkökul mezunu (n = 193, %22) babalar takip etmektedir. Bu durum veri setinde babaların eğitim düzeylerinin genellikle ortaöğretim ve üzeri seviyelerde yoğunlaştığını göstermektedir.

Tablo 6. Aile gelir durumu dağılımı
(Table 6. Distribution of family income status)

Kategori	Frekans	Yüzde (%)
Düşük	257	29.3
Orta	320	36.5
Yüksek	300	34.2
Toplam	877	100

Araştırmada veriler toplanırken ailelerinin gelir düzeyleri, asgari ücret temel alınarak üç kategoride toplanmıştır. Aylık geliri bir asgari ücretin altında olanlar düşük, aylık

geliri bir ile iki asgari ücret arasında olanlar orta ve aylık geliri iki asgari ücretin üzerinde olanlar yüksek gelir düzeyi olarak sınıflandırılmıştır. Bu sınıflamaya göre veri setindeki dağılım Tablo 6'da verilmiştir. Öğrencilerin en büyük bölümünü orta gelir düzeyindeki aileler oluşturmaktadır (n = 320, %36.5). Bunu sırasıyla yüksek gelir düzeyindeki (n = 300, %34.2) ve düşük gelir düzeyindeki (n = 257, %29.3) aile gelirleri takip etmektedir. Veri setinde öğrencilerin çoğunluğunun asgari ücretin üzerinde bir gelire sahip ailelerden geldiği görülmektedir.

Tablo 7. Katılımcıların mezun olduğu lise türü dağılımı
(Distribution of Participants by Type of High School Graduated)

Kategori	Frekans	Yüzde (%)
Anadolu Lisesi	371	42.3
Açık Öğretim Lisesi	6	0.7
Fen Lisesi	128	14.6
Güzel Sanatlar Lisesi	54	6.2
Meslek Lisesi	138	15.7
Sosyal Bilimler Lisesi	12	1.4
Spor Lisesi	67	7.7
İmam Hatip Lisesi	101	11.5
Toplam	877	100

Tablo 7'de 877 öğrencinin mezun oldukları lise türleri dağılımı verilmiştir. Tablo 7 incelendiğinde yüksek oran Anadolu Lisesi mezunlarına aittir (n = 371, %42.3). Bu oran, çalışmaya katılan her iki öğrenciden birinin Anadolu Lisesi kökenli olduğunu göstermektedir. Anadolu liseleri öğrencilerin yetenek, başarı ve ilgilerine göre öğrencilerin yükseköğretim programlarına hazırlanmalarını amaçlayan eğitim kurumlarıdır [45]. İkinci sırada yer alan Meslek Liseleri (n = 138, %15.7) ve üçüncü sıradaki Fen Liseleridir (n = 128, %14.6). Veri setindeki katılımcıların önemli bir kısmının teknik/mesleki ve fen bilimleri odaklı eğitim kurumlarından geldiğini göstermektedir. Bu durum, farklı lise türlerinden mezun öğrencilerin farklı akademik altyapılarla yükseköğretime geçtiklerini de ortaya koymaktadır.

Tablo 8. Öğrencilerin yerleştiği fakülteye göre dağılımı
(Table 8. Distribution of students by faculty placement)

Sıra No	Kategori	Frekans	Yüzde (%)
1.	Devlet Konservatuvarı	50	5.7
2.	Dış Hekimliği Fakültesi	67	7.7
3.	Eczacılık	52	5.9
4.	Eğitim Fakültesi	76	8.7
5.	Fen-Edebiyat Fakültesi	63	7.2
6.	Hukuk	56	6.4
7.	Mühendislik	55	6.3
8.	Sağlık Bilimleri Fakültesi	55	6.3
9.	Sosyal ve Beşeri Bilimler Fakültesi	51	5.8
10.	Spor Bilimleri	61	7
11.	Tıp Fakültesi	61	7
12.	Yabancı Diller	58	6.6
13.	Ziraat Fakültesi	58	6.6
14.	İktisadi ve İdari Bilimler	55	6.3
15.	İlahiyat	59	6.7
	Toplam	877	100

Tablo 8’de 877 öğrencinin yerleştikleri fakültelere göre dağılımı verilmiştir. Tablo 8 incelendiğinde veri setindeki en fazla öğrenci bilgisi Eğitim Fakültesi’ne yerleşen (n = 76, %8.8) öğrencilerden oluşmaktadır. Fakülteye göre dağılımda Eğitim Fakültesi’ni sırasıyla Diş Hekimliği Fakültesi (n = 67, %7.6), Fen-Edebiyat Fakültesi (n = 63, %7.2), Tıp Fakültesi ve Spor Bilimleri Fakültesi (n = 61, %7) izlemektedir. Bu fakülte dağılımlarına bakıldığında homojen ve hem sayısal hem de sözel ağırlıklı alanların veri setinde bulunduğu görülmektedir. Ayrıca Mühendislik, Sağlık Bilimleri ve İktisadi ve İdari Bilimler Fakülteleri gibi alanların eşit sayıda (n = 55, %6.3) olması uygulamalı ve disiplinler arası programlardan toplanan verinin dengeli dağıldığını görülmektedir. Bu bulgular, veri setinde bulunan öğrencilerin bilgileri yerleştiği fakülte çeşitliliği açısından homojen ve dengeli bir yapıya sahip olduğunu göstermektedir.

Tablo 2 Öğrencilerin ortaokul, lise ve üniversite başarı notları

(Table 9. Students’ success grades in middle school, high school, and university)

Değişken	Minimum	Maksimum	Ortalama
Ortaokul Başarı Notu	60	100	81.5
Lise Başarı Notu	63	100	82.8
Üniversite GNO	2.5	4,00	3

Tablo 9’da öğrencilerin akademik başarı notlarının okul düzeyine göre dağılımı verilmiştir. Tablo 10 incelendiğinde ortaokul ve lise notların 60 ile 100 arasında değiştiği görülmektedir. Üniversite genel not ortalamasının ise 3 olduğu görülmektedir. Üniversite öğrencilerinden veri toplanırken örneklemin 2.5 ve üstü not ortalaması olmasından dolayı genel not ortalaması (4 üzerinden 03) ve okul başarı not ortalamaları (ortaokul 81.5, lise 82.8) yüksektir. Bu durum üniversite başarısı ile okul başarı puanları arasında pozitif anlamlı bir ilişki olduğunu göstermektedir.

Tablo 10. Uygulanan sınıflandırma algoritmaları ve sonuçları

(Table 10. Applied classification algorithms and results)

Algoritma	Doğruluk (%)	Kappa Değeri	En Hatalı Sınıflar	En Başarılı Sınıflar
Random Forest	87.6	0.8665	9	3, 7, 10, 12-15
SMO (SVM)	85.4	0.8434	8, 9	3, 7, 10, 12-15
Naive Bayes	83.5	0.8226	9	3, 6, 7, 12, 15
J48	81.4	0.8008	4, 9	3, 7, 10, 12-15
IBk (k-NN)	79.4	0.7785	8,9,11	7, 10, 13-15

Veri setine lise öğrencilerin yerleştikleri fakültelerin tahmin edilmesine yönelik olarak 5 farklı sınıflandırma algoritmaları uygulanmış ve sonuçları Tablo 10’da verilmiştir. Bu algoritmaların doğruluk oranları, Cohen’s Kappa değeri ve model sonuç ekranındaki karmaşıklık matrisi temel alınarak, veri setindeki "yerleşilen fakülte" özneteliğine göre sınıflandırma doğrulukları analiz edilmiş; en başarılı ve en hatalı sınıflar belirlenmiştir. En başarılı ve en hatalı sınıflar sütununda bulunan numaralar Tablo 8’de verilen “Öğrencilerin yerleştiği fakülteye göre dağılımı” tablosunda fakülte sıra numaralarını ifade etmektedir.

Random Forest algoritması, %87.6 doğruluk oranı ve 0.8665 kappa değeri ile doğruluk oranı en yüksek olarak en başarılı sınıflandırma performansını sergilemiştir. Bu sonuç da sınıf tahminlerinin tesadüfen dağılmadığını ve modelin güvenilir olduğunu göstermektedir. Modelde en sorunlu sınıflandırma, “9 numara” olarak kodlanan Sosyal ve Beşerî Bilimler Fakültesi sınıfı olmuştur. Buna karşılık, en başarılı sınıflar ise 3 (Eczacılık), 7 (Mühendislik), 10 (Spor Bilimleri) ve 12–15 (Yabancı Diller, Ziraat Fakültesi, İktisadi ve İdari Bilimler, İlahiyat) olarak belirlenmiştir. Bu durum, Random Forest sınıflandırma etiketi çeşitliliği karşısında oldukça esnek ve güçlü olduğunu göstermektedir.

SMO (SVM) algoritması %85.4 doğruluk ve 0.8434 kappa değeriyle ikinci sırada yer almıştır. 8 (Sağlık Bilimleri Fakültesi) ve 9 (Sosyal ve Beşerî Bilimler Fakültesi) numara ile kodlanan sınıflar hata oranı yüksek iken, tıpkı Random Forest’ta olduğu gibi 3 (Eczacılık), 7 (Mühendislik), 10 (Spor Bilimleri) ve 12–15 (Yabancı Diller, Ziraat Fakültesi, İktisadi ve İdari Bilimler, İlahiyat) numara ile kodlanan sınıflar başarılı olmuştur.

Naive Bayes algoritması %83.5 doğruluk ve 0.8226 kappa değeriyle iyi bir genel performans sunmuştur. En hatalı sınıf yine 9 (Sosyal ve Beşerî Bilimler Fakültesi) numara ile kodlanan sınıf iken, 3 (Eczacılık), 6 (Hukuk), 7 (Mühendislik), 12 (Yabancı Diller) ve 15 (İlahiyat) numara ile kodlanan sınıfların başarısı öne çıkmaktadır.

J48 (Karar Ağacı) algoritması %81.4 doğruluk ve 0.8008 kappa değeriyle dördüncü sırada yer almaktadır. Modelde 4 (Eğitim Fakültesi) ve 9 (Sosyal ve Beşerî Bilimler Fakültesi) numara ile kodlanan sınıf etiketleri hatalı olarak belirlenmiş; buna karşılık 3 (Eczacılık Fakültesi), 7 (Mühendislik Fakültesi), 10 (Spor Bilimleri Fakültesi) ve 12–15 (Yabancı Diller Fakültesi, İktisadi ve İdari Bilimler Fakültesi, Ziraat Fakültesi, İlahiyat Fakültesi) numara ile kodlanan sınıflar yüksek performans göstermiştir.

IBk (k-En Yakın Komşu) algoritması ise %79.4 doğruluk oranı ve 0.7785 kappa değeriyle en düşük performansı göstermiştir. 8 (Sağlık Bilimleri Fakültesi), 9 (Sosyal ve Beşerî Bilimler Fakültesi) ve 11 (Tıp Fakültesi) numara ile kodlanan sınıf etiketleri en hatalılar iken; 7 Mühendislik, 10 Spor Bilimleri ve 13–15 (İktisadi ve İdari Bilimler Fakültesi, Ziraat Fakültesi, İlahiyat Fakültesi) numaraları ile kodlanan sınıf etiketleri başarılı olmuştur. Sınıflar arası

en yoğun karışıklık (8, 9, 11 kodlu sınıflar) bu modelde olduğu görülmüştür. Bu durumun giderilmesi için daha belirgin öznitelikler ile desteklenmeyi gerekli kılmaktadır.

Tablo 10'da yer alan modellerin Fen başarılı sınıfların doğru sınıflandırma oranları genel olarak değerlendirildiğinde; Eczacılık Fakültesi (3), Mühendislik Fakültesi (7), Spor Bilimleri Fakültesi (10), Yabancı Diller Fakültesi (12), Ziraat Fakültesi (13), İktisadi ve İdari Bilimler Fakültesi (14) ve İlahiyat Fakültesi (15) fakültelerinin tüm algoritmalar tarafından yüksek doğrulukla sınıflandırıldığı görülmektedir. Buna karşın 9 (Sosyal ve Beşeri Bilimler Fakültesi) numara ile kodlanan sınıfın çoğu modelde sınıflandırması hatalı olduğu görülmektedir. Bu durum, söz konusu sınıfların sahip olduğu öznitelik desenlerinin ayırt edici nitelikte olduğunu göstermektedir. Örnek verecek olursak bu fakültelere yerleşen öğrencilerin mesleki kişilikleri, kişilik profilleri, akademik başarı düzeyleri gibi özniteliklerin diğer fakültelerden belirgin şekilde farklılaşmış olabilir.

Tablo 10'da yer alan sınıflandırma algoritmaları modellerinden elde edilen kappa değerleri 0.77 ile 0.87 arasında değiştiği görülmektedir. Kappa değerinin bu aralıkta olması sınıflandırıcıların tesadüfi başarıyı önemli ölçüde aştığını ve modellerin güvenilir tahminler üretebildiğini göstermektedir [46].

Fakülte yerleştirmede en etkili öznitelikleri belirlemek için başarı performansı en yüksek çıkan Random Forest algoritması modelinin öznitelik önemi analizi (Feature Importance Analysis) yapılarak öznitelik önem puanları çıkarılmıştır. Bu kapsamda en belirleyici 10 öznitelik Tablo 11'de verilmiştir.

Tablo 11. Random Forest modelinin öznitelik önem analizi sonucu (ilk 10 öznitelik)

(Table 11. Feature importance analysis result of the Random Forest model (top 10 features))

Sıra No.	Öznitelik	Açıklaması	OSA	Node
1	Baba_e_duz	Baba eğitim düzeyi	0.72	236
2	Anne_e_duz	Anne eğitim düzeyi	0.71	244
3	Yas	Yaş	0.68	372
4	Cinsiyet	Cinsiyet	0.67	96
5	Aile_gelir	Aile gelir durumu	0.66	170
6	RIASEC_Gercekci_1	Elektronik aletleri tamir etmeyi severim.	0.62	312
7	RIASEC_Gercekci_2	Evde bozulan eşyaları tamir etmekten hoşlanırım.	0.62	340

8	RIASEC_Gercekci_3	Araba tamiriyle ilgilenirim.	0.61	319
9	lise_basari	Lise mezuniyet başarı ortalaması.	0.59	424
10	RIASEC_Gercekci_5	İş makinelerinin nasıl kullanıldığını merak ederim.	0.59	296

Tablo 11'de öznitelikler, açıklamaları, "önem" değeri olarak ortalama safsızlık azalması (OSA) (average impurity decrease) ve özniteliğin kullanıldığı düğüm (node) sayısını verilmiştir. Ortalama safsızlık azalması (average impurity decrease), karar ağaçlarında modellerde kullanılan öznitelik önemini ölçme yöntemlerinden biridir. Bu yöntem, bir özniteliğin model içindeki karar düğümlerinde kullanıldığında sağladığı bilgi kazancı veya "safsızlıkta" (impurity) meydana gelen azalma üzerinden değerlendirilir [40, 47, 48]. Modelde karar verme sürecine katkısını ifade eden nicel bir değerdir. Bu nedenle ortalama safsızlık azalması yüksek olan öznitelikler, karar verme sürecinde daha belirleyici bir rol üstlenmektedir.

Tablo 11 incelendiğinde ebeveyn eğitim düzeyi çok yüksek ortalama safsızlık azalması değerine sahiptir (anne eğitim düzeyi = 0.72, baba eğitim düzeyi = 0.71). Bu durum öğrencinin yerleştiği fakülte üzerinde ailenin eğitim düzeyi durumunun belirleyici olduğunu göstermektedir. Aile gelir özniteliği de karar sürecinde etkili olduğu görülmüştür. Demografik faktörlerden yaş, cinsiyet ve sınıf düzeyi, sınıflandırma tahmini doğruluğunu doğrudan etkilemektedir. Bu değişkenler fakülte tercihlerinde etkili olduğunu göstermektedir. Lise başarı performansı da modelin tahmin başarısında güçlü özniteliklerinden biridir. Mesleki kişilik envanteri maddeleri de modelin tahmininin de ön plana çıkan öznitelikler arasındadır. Bu durum mesleki kişiliğin fakülte tercihlerinde etkili olduğunu göstermektedir.

Random Forest algoritmasıyla gerçekleştirilen öznitelik önem analizi sonucunda, her iki ölçekte yer alan maddelerin fakülte tahmininde farklı düzeylerde etkili olduğu görülmüştür. BFKÖ ölçeğinde yer alan maddelerin OSA değerleri 0.36 ile 0.47 arasında değişmektedir. BFKÖ ölçeği'ne ait "Kendimi kolay sınırlanan biri olarak görüyorum." (OSA=0.36) ve "Kendimi dışa dönük, sosyal biri olarak görüyorum." (OSA=0.36) maddeleri modelde en düşük belirleyicilik düzeyine sahip öznitelikler olarak bulunmuştur. Buna karşılık "Kendimi bir işi tam yapacak biri olarak görüyorum" (OSA=0.47) maddesi, BFKÖ ölçeği içinde en yüksek ama genel sonuca orta düzeyde belirleyici bir öznitelik olarak öne çıkmıştır. Genel olarak kişilik özelliklerinin fakülte yerleşiminde yüksek bir etki göstermediği ve dolaylı bir rol oynadığı söylenebilir. RIASEC Mesleki Kişilik Tipleri ölçeğinde ise OSA değerlerinin 0.47 ile 0.62 arasında değiştiği görülmüştür. RIASEC Mesleki Kişilik Tipleri Envanteri'ne ait "Bir müzik aleti çalmak isterim" (OSA=0.47) maddesi, ölçek

içinde en düşük etki düzeyine sahip madde olarak belirlenmiştir. Buna karşın, aynı ölçek kapsamında yer alan “Elektronik aletleri tamir etmeyi severim” (OSA=0.62) ve “Evde bozulan eşyaları tamir etmekten hoşlanırım” (OSA=0.62) maddeleri hem ölçek kendi içinde hem de genel sonuca yüksek düzeyde belirleyici öznetelik olarak bulunmuştur. Bu bulgu, öğrencilerin fakülte tercihlerinde bireyin mesleki ilgi yönelimlerinin kişilik özelliklerinden daha belirleyici olduğunu ortaya koymaktadır.

Tablo 12. Makine Öğrenmesi Algoritmalarının Performans Ölçütleri ve Hata Metrikleri Karşılaştırması
(Table 12. Comparison of performance and error metrics of machine learning algorithms)

Algoritma	F1-Skoru	Kesinlik	Duyarlılık	Ortalama Mutlak Hata	Kök Ortalama Kare Hata
Random Forest	0.867	0.901	0.876	0.0314	0.1142
Support Vector Machine (SVM)	0.850	0.850	0.854	0.1159	0.2351
Naive Bayes	0.823	0.835	0.835	0.0229	0.1360
Decision Tree (J48)	0.812	0.811	0.814	0.0259	0.1503
IBk (k-NN)	0.792	0.796	0.794	0.0293	0.1644

Tablo 12’de görüldüğü üzere makine öğrenmesi algoritmalarının performans ölçütleri ve hata metrikleri birlikte verilmiştir. Elde edilen bulgular, Random Forest algoritmasının diğer modellere kıyasla en yüksek performansı gösterdiğini ortaya koymaktadır (F1-skoru = 0.867; kesinlik = 0.901; duyarlılık = 0.876). Bu sonuç, topluluk (ensemble) öğrenme yaklaşımının, özellikle sınıflandırma doğruluğu açısından istikrarlı ve genellenebilir sonuçlar ürettiğini göstermektedir. Random Forest’ın yüksek kesinlik değeri, yanlış pozitiflerin (false positive) düşük olduğunu; yüksek duyarlılık değeri ise modelin doğru sınıfları yakalama yeteneğinin güçlü olduğunu göstermektedir.

İkinci en başarılı model, Support Vector Machine (SVM) olmuştur (F1-skoru = 0.850). SVM’nin özellikle karmaşık karar sınırlarını tanımlamada güçlü olduğu bilinmektedir; ancak yüksek ortalama mutlak hata (0.1159) ve kök ortalama kare hata (0.2351) değerleri, modelin bazı örneklerde hatalı sınıflandırmalar ürettiğini göstermektedir.

Naive Bayes modeli, basit varsayımlarına rağmen tatmin edici bir performans sergilemiştir (F1-skoru = 0.823). Bu durum, değişkenler arasındaki bağımsızlık varsayımının veri setinde ciddi bir performans kaybına yol açmadığını göstermektedir. Dikkat çekici biçimde, Naive Bayes modelinin ortalama mutlak hata değeri (0.0229) oldukça düşük çıkmıştır; bu da hata dağılımının dengeli olduğunu işaret etmektedir.

Decision Tree (J48) ve IBk (k-NN) algoritmaları ise görece düşük F1-skorlarına sahiptir (sırasıyla 0.812 ve 0.792). Bu durum, özellikle karmaşık ve yüksek boyutlu veri

yapılarında tek ağaç tabanlı modellerin ve örnek temelli yaklaşımların genelleme kapasitesinin sınırlı olmasından kaynaklanabilir.

Hata metrikleri incelendiğinde, en düşük ortalama mutlak hata ve kök ortalama kare hata değerlerinin yine Random Forest modeline ait olduğu görülmektedir (sırasıyla 0.0314 ve 0.1142). Bu sonuç, modelin tahmin hatalarının küçük ve dengeli dağıldığını göstermektedir. Buna karşılık, SVM modelinde hem ortalama mutlak hata hem de kök ortalama kare hata değerleri yüksek çıkmıştır; bu da modelin bazı örneklerde ciddi sapmalar üreteceği anlamına gelebilir.

Genel olarak Random Forest modeline göre fakülte yerleşiminde en belirleyici değişkenler; öğrencinin ailesine, akademik geçmişine ve mesleki kişilik tiplerine dayanmaktadır. Bu bulgu, kariyer danışmanlığı ve fakülte yerleşim öngörülerinde bu tür değişkenlerin dikkate alınması gerektiğini göstermektedir.

4. SONUÇ VE TARTIŞMA (Results and Discussion)

Bu çalışmada, makine öğrenmesi algoritmaları ve veri madenciliği yöntemlerinden yararlanarak lise düzeyindeki öğrencilerin mesleki kişilik, kişilik ve başarı özelliklerine dayanarak hangi fakültede başarılı olabilecekleri tahmininde bulunan bir sınıflandırma modeli geliştirilmiştir. Araştırmada 877 öğrenciden 51 öznetelikte elde edilen veri seti, makine öğrenmesi temelli beş farklı sınıflandırma algoritması ile analiz edilerek modeller oluşturulmuştur. Kullanılan sınıflandırma algoritmaları: Random Forest, SMO (SVM), Naive Bayes, J48 (C4.5) ve IBk (k-Nearest Neighbors)’dır. Veri setinde sınıf etiketi olarak öğrencinin öğrenim gördüğü fakülte bilgisi esas alınmıştır. Oluşan modellerin performansları; doğruluk oranı, Kappa değeri ve karmaşıklık matrisleri karşılaştırılarak değerlendirilmiştir. Bu değerlendirme, sonuçların daha tutarlı bir şekilde değerlendirerek sınıf bazında modelin güçlü ve zayıf yönlerini ortaya koymak ve yorumlamak hedeflenmiştir.

Veri seti üzerinde yapılan analizler sonucunda, Random Forest algoritması ile oluşan modelin %87.6 doğruluk oranı ve 0.8665 kappa katsayısı ile oluşan beş model içinde en başarılı sınıflandırma modeli olmuştur. Bu algoritmadan oluşan modeli sırasıyla SMO (%85.4 doğruluk, $\kappa=0,8434$), Naive Bayes (%83.47 doğruluk, $\kappa=0,8226$), J48 (%81.4 doğruluk, $\kappa=0,8008$) ve IBk (%79.4 doğruluk, $\kappa=0,7785$) sınıflandırma algoritmalarından oluşan modeller takip etmiştir. Bu bulgu, literatürde sınıflandırma algoritmalarının performansını ölçen birçok araştırmada Random Forest algoritması ile çıkan model başarılı bulunmuştur [49, 50, 51, 52, 53]. Benzer şekilde [54] tarafından yapılan bir çalışmada, Random Forest algoritması ile öğrencilerin ilk iki dönem notlarına dayanarak lisans derecesi alıp alamayacağı ve kaydoldukları birkaç derse bakılarak ana bölümleri tahmin edilmiştir. Söz konusu çalışma Random Forest algoritmasını bu tür tahminlerde etkin kullanılabileceğini göstermektedir.

Araştırmada kullanılan sınıflandırma algoritmalarının karmaşıklık matrisleri en yüksek ve en hatalı sınıfların belirlenmesi amacıyla karşılaştırmalı olarak incelendiğinde; Tıp, Mühendislik, Spor Bilimleri, Yabancı Diller, Ziraat, İktisadi ve İdari Bilimler ile İlahiyat fakülteleri tüm modellerde yüksek doğrulukla ile sınıflandırma işlemi yapılmıştır. Buna karşılık, Sosyal ve Beşerî Bilimler Fakültesi ortaya çıkan modellerin hepsinde en düşük doğru sınıflandırma oranına sahip sınıf olmuştur. Bu durum bu fakülteye yerleşen öğrencilerin özellikleri diğer fakültelere yerleşen öğrencilerin özelliklerine yakın olduğu ve bundan dolayı ayırım yapmakta zorlandığını göstermektedir. Bu sınıfın içerisinde Bilgi ve Belge Yönetimi, Türk Dili ve Edebiyatı, İletişim ve Tasarımı gibi çeşitli bölümlerdeki öğrenciler yer almaktadır. Dolayısıyla bu sınıftaki öğrencilerin kişilik özelliklerinin çeşitlilik göstermesi de beklendik bir sonuçtur. Ek olarak, sınıflandırma modellerinin başarısının düşmesinde “etiket gürültüsü” (label noise) önemli bir rol oynamaktadır. [55]’in yaptığı kapsamlı inceleme, etiket hatalarının hem sınıflandırma doğruluğunu düşürdüğünü hem de model karmaşıklığını artırdığını vurgulamaktadır. Sosyal ve Beşerî Bilimler fakültesinde heterojenlikle birlikte etiket gürültüsünün daha etkili olması, bu sınıfın diğerlerine göre istikrarlı biçimde daha düşük doğrulukta sınıflandırılmasına yol açabilir. Makine öğrenmesi tabanlı “misclassification network” (yanlış sınıflandırma ağı) yöntemi, sosyal bilimler öğrencilerinin özelliklerinin hangi fakültelere daha yakın olduğunu görselleştirmek için kullanılabilir. [55], yanlış sınıflandırma ağlarının bilimsel disiplinler arasındaki örtüşmeleri ortaya çıkarabildiğini göstermiştir.

Random Forest algoritmasıyla oluşturulan modelin öznitelik önem analizi sonuçlarına göre, sınıf etiketi olan yerleşilen fakülte öznitelğine en yüksek etki eden ilk on öznitelik incelendiğinde demografik özellikler bakımından baba eğitim düzeyi, anne eğitim düzeyi, yaş, cinsiyet ve aile gelir durumu gibi değişkenlerin belirleyici olduğu görülmüştür. Bu bulgu literatürdeki çalışmalarla benzerlik taşımaktadır. Araştırmalar öğrencilerin yaş, cinsiyet, anne-baba eğitim durumu ve sosyo-ekonomik düzeyine göre meslek seçimlerinin anlamlı bir şekilde farklılaştığını ortaya koymaktadır [57, 58, 59, 60]. [26] tarafından yapılan araştırmada öğrencilerin kariyer seçiminde cinsiyet, ebeveyn eğitim durumu ve ekonomik gelir düzeyinin etkili olduğu sonucu edilmiştir. Mevcut araştırmada akademik başarı bağlamında lise mezuniyet başarı ortalamasının yerleşilen fakülte öznitelğinde belirleyici olduğu bulgusu elde edilmiştir. Ülkemizde lise türleri başarı olarak farklılaşmaktadır ve mezun olunan lise türü, üniversitede devam edilecek bölümün seçimini etkilemektedir. [61] tarafından yapılan araştırmada fen lisesi mezunu öğrencilerin tıp, eczacılık, diş hekimliği gibi sağlık alanlarını tercih ettikleri, genel lise mezunlarının ise eğitim ve fen edebiyat fakültelerini tercih ettikleri sonucunu elde etmişlerdir.

Bunun yanı sıra Mesleki Kişilik Envanteri kapsamında “Elektronik aletleri tamir etmeyi severim”, “Evde bozulan eşyaları tamir etmekten hoşlanırım”, “Araba tamiriyle

ilgilienirim.” ve “İş makinelerinin nasıl kullanıldığını merak ederim.” ifadelerinin fakülte seçiminde belirleyici olduğu saptanmıştır. Bu öznitelikler model üzerinde en yüksek etkiye sahip olduğu görülmektedir. Bu bulgu, bireylerin kişilik özellikleriyle uyumlu mesleklere yönelmelerinin hem bireysel gelişim hem de toplumsal fayda açısından önemli katkılar sağladığını göstermektedir. [7] çalışmasında öğrencilerin kişilik özellikleri ile kariyer tercihleri arasında uyumsuzluk belirlediğini, bu noktada köklü kariyer danışmanlığı çalışmalarına ihtiyaç duyduklarına dikkat çekmektedir. Ayrıca literatürde, kişilik özelliklerinin öğrencilerin akademik başarıları üzerinde etkili olduğu belirtilmektedir [62]. Bu bağlamda kariyer danışmanlığı sürecinde öğrencilerin kişilik özelliklerinin dikkate alınarak yönlendirme yapılması gerekmektedir. Mevcut araştırma sonucunda yerleşilen fakültede kişilik özelliklerinin yüksek etkiye sahip olduğu belirlenmiştir. Literatür incelendiğinde araştırma sonucunun benzer çalışmalarla paralel olduğu görülmüştür [63, 64, 1]. Bu sonuç, veri temelli kariyer rehberliği sistemlerinde kişilik ve ilgi envanterlerinin dikkate alınmasının önemini bir kez daha ortaya koymaktadır. Ayrıca, aile eğitimi ve gelir düzeyinin de yerleşilen fakülte üzerinde etkili olması, sosyoekonomik faktörlerin yükseköğretim yönelimi üzerindeki belirleyici rolünü teyit etmektedir. Gelecek çalışmalarda, bu özniteliklerin zaman içerisindeki değişimi ve farklı okul türleri arasındaki etkileri de incelenebilir.

Çalışmada kullanılan BFKÖ ölçeği ve RIASEC Mesleki Kişilik Tipleri Envanteri’nin tüm maddeleri modelin sınıf etiketini belirlemede değerlendirilmiş; elde edilen sonuçlar, en etkili özniteliklerin RIASEC Mesleki Kişilik Tipleri Envanteri’ne ait maddelerden oluştuğunu göstermiştir. Bu bulgu, öğrencilerin fakülte yerleşiminde mesleki ilgi ve yönelimlerinin kişilik özelliklerine kıyasla daha yüksek ayırt edici güce sahip olduğunu ortaya koymaktadır. Araştırma bulgusu benzer araştırma sonuçları ile paralel niteliktedir. Lise öğrencilerinin RIASEC model (ilgi yönelimleri) ve Big Five personality traits (kültürde kişilik özellikleri) karşılaştırıldığı araştırmada, öğrencilerin araştırmacı veya girişimci ilgileri yüksek olan öğrencilerin daha çok genel eğitim alanını seçerken, sosyal veya geleneksel ilgileri yüksek olan öğrencilerin ise daha çok mesleki alanı tercih ettiği sonucuna ulaşılmıştır. Kişilik özelliklerinin ise eğitim alanı seçimleriyle hiçbir anlamlı ilişkisi olmadığı veya çok az anlamlı ilişkisi olduğu görülmüştür [65]. Benzer olarak 3 bin kişi üzerinde 10 yıllık izleme çalışmasında lise sonundaki RIASEC ilgi yönelimleri ile Big Five kişilik ölçümleri karşılaştırılmıştır. Araştırma sonucunda öğrencilerin iş ve ilişki çıktılarında, mesleki ilgi alanlarının kişilik özelliklerinden daha güçlü yordayıcılar olduğu görülmüştür [66]. Bu bağlamda, öğrencilerin ilgi yönelimleri, kişilik özelliklerine kıyasla eğitim-meslek seçiminde daha güçlü bir belirleyici olabilir.

Tüm bu bulgulara bakıldığında Random Forest algoritması modeli yerleşilen fakülte tahmininde en yüksek performansı göstermiş olup öğrencinin akademik başarıları kadar mesleki kişilik ve demografik özelliklerinin de

öğrencinin fakülteadaki başarısı açısından önemli etkenler olduğunu göstermektedir. Bu sonuçlar, veriye dayalı kariyer yönlendirme sistemlerinin geliştirilmesi açısından önemli bir referans niteliğindedir. Ek olarak gelecek çalışmalarda, bu çalışmada geliştirilen makine öğrenmesi tabanlı modelin kullanıcı ara yüzü içeren işlevsel bir sistem olarak uygulanabilir hale getirilmesi ve farklı disiplinlerdeki öğrencilere yönelik uygulanabilirliğinin test edilmesi önerilmektedir.

Son olarak bu araştırmanın bazı sınırlılıkları mevcuttur. Bu araştırma Tokat Gaziosmanpaşa Üniversitesi ve Amasya Üniversitesi'nde öğrenim görmekte olan üniversite öğrencileri üzerinde gerçekleştirilmiştir. Bu bağlamda araştırma sonuçlarının farklı üniversitelerde öğrenim görmekte olan öğrenciler için geçerliliğini artırmak amacıyla daha çeşitli üniversitelerde uygulanması önerilebilir. İkincisi bu çalışmada öğrencilerin kişilik özellikleri ve başarı durumları öz bildirim tarzı ölçekler aracılığıyla incelenmiştir. Bu nedenle araştırma sonuçları öz bildirim tarzı ölçeklere karışan sosyal beğenilirlik, orta yol cevap verme gibi hataları içermektedir [67]. Bu değişkenlerin yanı sıra gelecek çalışmalarda, kariyer seçimi için önemli olan yetenek, ilgi gibi değişkenlerin de ele alınması önerilebilir.

KAYNAKLAR (REFERENCES)

- [1] E. Yanikkerem, S. Altınparmak, G. Karadeniz, "Gençlerin meslek seçimini etkileyen faktörler ve benlik saygıları", *Nursing Forum Dergisi*, 7(2), 61–62, 2004.
- [2] R. L. Atkinson, R. C. Atkinson, E. E. Smith, D. J. Bem, S. Hoeksema, **Psikolojiye Giriş**, Arkadaş Yayınları, İstanbul, 1999.
- [3] E. Köroğlu, S. Bayraktar, **Kişilik Bozuklukları**, HYB Basım Yayın, Ankara, 2011.
- [4] J. M. Burger, **Kişilik: Psikoloji Biliminin İnsan Doğasına Dair Söyledikleri**, İ. D. Ergüven Sarıoğlu (Çev.), Kaknüs Yayınları, İstanbul, 2006.
- [5] R. R. McCrae, P. T. Costa, "Four ways five factors are basic", *Personality and Individual Differences*, 13(6), 653–665, 1992.
- [6] L. R. Goldberg, "An alternative "description of personality": The Big-Five factor structure", *Journal of Personality and Social Psychology*, 59, 1216–1229, 1990.
- [7] V. Cosmas, "Personality traits and career preferences among secondary school students in Tanzania: A match within introvert-extrovert and agreeable-less-agreeable traits", **Proceedings of the 1st International Conference of Education**, 2023.
- [8] Ö. Söner, O. Yılmaz, "How do personality traits influence adolescents' career anxiety and self-efficacy in making career decisions?" *Journal of Teacher Education and Lifelong Learning*, 4(2), 110–123, 2022.
- [9] Ö. Üfelek, E. Yavuz, "Kişilik tipleri ile kariyer algısı ve seçimi arasındaki ilişkinin incelenmesi". *Journal of Recreation and Tourism Research*, 8(3), 289–313, 2021.
- [10] S. Yüksel, E. Polat, S. Sağ, "Öğrencilerde kişilik tipleri ve kariyer kararı: spor bilimleri ve diğer bölüm öğrencilerinin özellikleri üzerine bir inceleme." *Uluslararası Bozok Spor Bilimleri Dergisi*, 5(3), 286–298, 2024.
- [11] F. Bacanlı, "Özellik-Faktör Uyumlu Kuramlar", In B. Yeşilyaprak (Ed.), *Mesleki Rehberlik ve Kariyer Danışmanlığı: Kuramdan Uygulamaya*, Pegem Akademi, Ankara, 2014.
- [12] J. Rayman, L. Atanasoff, "Holland's theory and career intervention: The power of the hexagon", *Journal of Vocational Behavior*, 55(1), 114–126, 1999.
- [13] D. M. Dozier, B.-L. Sha, H. Shen, "Why women earn less than men: The cost of gender discrimination in U.S. public relations", *Public Relations Journal*, 7(1), 1–21, 2013.
- [14] J. L. Holland, "Exploring careers with a typology: What we have learned and some new directions", *American Psychologist*, 51(4), 397–406, 1996.
- [15] R. S. Sharf, *Applying Career Development Theory to Counseling*, 6. cilt, Brooks/Cole, Belmont, CA, 2013.
- [16] S. G. Niles, J. Harris-Bowlsbey, 21. Yüzyılda Kariyer Gelişimi Müdahaleleri, F. K. Owen (Çev.), Nobel Akademik, Ankara, 2013.
- [17] M. M. Nauta, "Holland's Theory of Vocational Choice and Adjustment", In S. D. Brown & R. W. Lent (Eds.), *Career Development and Counseling: Putting Theory and Research to Work*, 2. cilt, John Wiley & Sons, Hoboken, NJ, 55–82, 2013.
- [18] B. Yeşilyaprak, 21. Yüzyılda Eğitimde Rehberlik Hizmetleri, Nobel Akademik, Ankara, 2013.
- [19] H. Al-Sahaf, Y. Bi, Q. Chen, A. Lensen, Y. Mei, Y. Sun, B. Tran, B. Xue, M. Zhang, "A survey on evolutionary machine learning", *Journal of the Royal Society of New Zealand*, 49, 205–228, 2019.
- [20] M. Jordan, T. Mitchell, "Machine learning: Trends, perspectives, and prospects", *Science*, 349, 255–260, 2015.
- [21] D. Lary, A. Alavi, A. Gandomi, A. Walker, "Machine learning in geosciences and remote sensing", *Geoscience Frontiers*, 7, 3–10, 2016.
- [22] Q. C. Song, H. J. Shin, C. Tang, A. Hanna, T. Behrend, "Investigating machine learning's capacity to enhance the prediction of career choices" *Personnel Psychology*, 77(2), 295–319, 2022.
- [23] P. Bahalkar, P. Peddi, S. Jain, "AI-driven career guidance system: A predictive model for student subject recommendations based on academic performance and aspirations", *Frontiers in Health Informatics*, 13(3), 2676–7104, 2024.
- [24] P. Shahane, P. Rinke, T. Datar, S. Badjate, "Student's career interest prediction using machine learning" *International Research Journal of Engineering and Technology (IRJET)*, 9(11), 533–539, 2022.
- [25] G. Akçapınar, E. Coşgun, "Öğrencilerin STEM eğitimi tercihlerinin veri madenciliği yaklaşımı ile tahmin edilmesi", *Eğitim Teknolojisi Kuram ve Uygulama*, 9(1), 73–88, 2019.
- [26] E. Özeren, T. Çiloğlu, R. Yılmaz, A. Özeren, "Öğrencilerin akademik kariyer hedefi seçiminde etkili olan faktörlerin veri madenciliği yöntemi ile belirlenmesi: Bartın başarı takip araştırması sonuçları üzerine bir inceleme", *Bilgi ve İletişim Teknolojileri Dergisi*, 2(2), 182–210, 2020.

- [27] E. Altunbaş, C. Altıntaş “Bilgisayar bölümü öğrencileri için makine öğrenme algoritmaları ile kariyer önerisi ve açıklanabilir yapay zeka ile değerlendirilmesi”, *Uluslararası Sürdürülebilir Mühendislik ve Teknoloji Dergisi*, 9(1), 133–149, 2025
- [28] D. Hand, H. Mannila, P. Smyth, “Principles of data mining”, *Drug Safety*, 30, 621–622, 2001.
- [29] N. Chawla, “Verilerde bilgiyi keşfetmek: Veri madenciliğine giriş”, *Biyoenformatikte Brifingler*, 6, 411–412, 2005.
- [30] H. Akpınar, “Veritabanlarında bilgi keşfi ve veri madenciliği”, *İ.Ü. İşletme Fakültesi Dergisi*, 29(1), 1–22, 2000.
- [31] Ş. Büyüköztürk, E. K. Çakmak, Ö. E. Akgün, Ş. Karadeniz, F. Demirel, *Bilimsel Araştırma Yöntemleri*, 17. baskı, Pegem Akademi Yayıncılık, Ankara, 2014.
- [32] T. M. Mitchell, *Machine Learning*, McGraw-Hill, New York, 1997.
- [33] M. B. Horzum, T. Ayas, M. A. Padır, “Beş Faktör Kişilik Ölçeğinin Türk Kültürüne Uyarlanması”, *Sakarya University Journal of Education*, 7(2), 398–408, 2017.
- [34] A. Atlı, G. Keldal, “Mesleki kişilik tipleri envanterinin geliştirilmesi”, *Eskişehir Osmangazi Üniversitesi Sosyal Bilimler Dergisi*, 18(1), 73–93, 2017.
- [35] B. Rammstedt, O. P. John, “Measuring personality in one minute or less: A 10-item short version of the Big Five Inventory in English and German”, *Journal of Research in Personality*, 41, 203–212, 2007.
- [36] H. W. Marsh, K. T. Hau, C. Artelt, J. Baumert, J. L. Peschar, “OECD’s brief self-report measure of educational psychology’s most useful affective constructs: Cross-cultural, psychometric comparisons across 25 countries”, *International Journal of Testing*, 6(4), 311–360, 2006.
- [37] K. Schermelleh-Engel, H. Moosbrugger, H. Müller, “Evaluating the fit of structural equation models: Tests of significance and descriptive goodness-of-fit measures”, *Methods of Psychological Research Online*, 8(2), 23–74, 2003.
- [38] I. H. Witten, E. Frank, M. A. Hall, *Data mining: Practical machine learning tools and techniques*, 3rd ed., Morgan Kaufmann Publishers, Burlington, MA, 2011.
- [39] J. R. Quinlan, *C4.5: Programs for Machine Learning*, Morgan Kaufmann Publishers, San Mateo, CA, 1993.
- [40] L. Breiman, “Random forests”, *Machine Learning*, 45(1), 5–32, 2001.
- [41] G. I. Webb, “Naïve Bayes”, In C. Sammut & G. I. Webb (Eds.), *Encyclopedia of Machine Learning and Data Mining*, Springer, 2017.
- [42] I. H. Witten, E. Frank, M. A. Hall, C. J. Pal, *Data Mining: Practical Machine Learning Tools and Techniques*, 4. cilt, Morgan Kaufmann Publishers, Cambridge, MA, 2016.
- [43] J. C. Platt, *Sequential Minimal Optimization: A Fast Algorithm for Training Support Vector Machines*, Technical Report MSR-TR-98-14, Microsoft Research, 1998.
- [44] M. Heydarian, T. Doyle, R. Samavi, “MLCM: Multi-label confusion matrix”, *IEEE Access*, 10, 19083–19095, 2022.
- [45] Internet: T.C. Millî Eğitim Bakanlığı, Okul Türleri, https://ogm.meb.gov.tr/meb_iys_dosyalar/2021_12/09162636_Okul_Turleri.pdf, 01.06.2025.
- [46] J. R. Landis, G. G. Koch, “The measurement of observer agreement for categorical data”, *Biometrics*, 33(1), 159–174, 1977.
- [47] L. Breiman, *Manual on Setting Up, Using, and Understanding Random Forests V3.1*, Statistics Department, University of California, Berkeley, CA, 2002.
- [48] G. Louppe, L. Wehenkel, A. Suter, P. Geurts, “Understanding variable importances in forests of randomized trees”, In *Advances in Neural Information Processing Systems*, Vol. 26, pp. 431–439, 2013.
- [49] C. Beaulac, J. S. Rosenthal, “Predicting university students’ academic success and major using random forests”, *arXiv preprint arXiv:1802.03418v3*, 15 January 2019.
- [50] C. U. Bertrand, O. B. Aliche, C. G. Onukwugba, C. I. Ofoegbu, D. A. Kelechi, I. C. Ugbor, N. M. Oragba, “Career guidance system using decision tree, random forest, and naïve Bayes algorithm”, *International Journal of Science, Technology and Society*, 13(2), 35–42, 2025.
- [51] S. Çınaroğlu, K. Avcı, “Dünya Sağlık Örgütü’ne üye ülkelerin doğumda beklenen yaşam süresi bakımından veri madenciliği yöntemleri kullanılarak sınıflandırılması”, *Istanbul University Econometrics and Statistics e-Journal*, (20), 88–103, 2014.
- [52] N. Karakurt, H. Özkan, “Öğrencilerin akademik başarısını etkileyen demografik değişkenlerin eğitsel veri madenciliği ile incelenmesi”, *Elektronik Eğitim Bilimleri Dergisi*, 13(25), 16–33, 2024.
- [53] B. Y. Korkmaz, U. Özcan, “Bir havayolu işletmesinde kabin ekiplerinin performans değerlendirmesine ilişkin bir veri madenciliği uygulaması”, *Karadeniz Fen Bilimleri Dergisi*, 8(2), 154–170, 2018.
- [54] C. Beaulac, J. S. Rosenthal, “Predicting university students’ academic success and major using random forests”, *arXiv preprint arXiv:1802.03418v3*, 15 January 2019.
- [55] B. Frénay, M. Verleysen, “Classification in the presence of label noise: A survey”, *IEEE Transactions on Neural Networks and Learning Systems*, 25(5), 845–869, 2014.
- [56] A. Lyutov, Y. Uygun, M. T. Hütt, “Machine learning misclassification of academic publications reveals non-trivial interdependencies of scientific disciplines”, *Scientometrics*, 126, 1173–1186, 2021.
- [57] B. E. Çapan, F. Korkut Owen, “Çocukların kendilik değerini geliştirmede kendilik değeri geliştirme programı ve sosyal ve duygusal eğitim programının etkililiği”, *Elektronik Sosyal Bilimler Dergisi*, 16(62), 756–769, 2017.
- [58] A. Soysal, K. Arıcıoğulları, “Meslek seçiminde yeterlilik ve meslek seçimini etkileyen faktörler: Sağlık yönetimi öğrencileri üzerinde bir araştırma”, *Kahramanmaraş Sütçü İmam Üniversitesi Sosyal Bilimler Dergisi*, 18(3), 2138–2153, 2021.
- [59] F. Vurucu, **Meslek Lisesi Öğrencilerinin Meslek Seçimi Yeterliliği ve Meslek Seçimini Etkileyen Faktörler**, Yüksek Lisans Tezi, Yeditepe Üniversitesi, Sosyal Bilimler Enstitüsü, İstanbul, 2010.

- [60] İ. A. Yılmaz, B. Dursun, K. Pektaş, A. Altay, “Üniversite öğrencilerinin kariyer seçimlerinin demografik özellikler açısından incelenmesi: Pınarhisar MYO örneği”, *Electronic Journal of Vocational Colleges*, 2(2), 9–21, 2012.
- [61] Y. Z. Ayık, A. Özdemir, U. Yavuz, “Lise türü ve mezuniyet başarısının kazanılan fakülte ile ilişkisinin veri madenciliği tekniği ile analizi”, *Atatürk Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, 10(2), 441–454, 2007.
- [62] J. R. Rico-Juan, C. Cachero, H. Macià, “Study regarding the influence of a student’s personality and an LMS usage profile on learning performance using machine learning techniques”, *Applied Intelligence*, 54, 6175–6197, 2024.
- [63] B. Erdem, M. F. Kayran, “Balıkesir Üniversitesi Turizm İşletmeciliği ve Otelcilik Yüksekokulu öğrencilerinin meslek seçimini etkileyen faktörler üzerine bir araştırma”, *Cumhuriyet Üniversitesi İktisadi ve İdari Bilimler Dergisi*, 14(1), 81–106, 2013.
- [64] Ö. Köroğlu, “Meslek seçimi ile kişilik özellikleri arasında ilişkinin belirlenmesi: Turizm rehberliği öğrencileri üzerine bir araştırma”, *Süleyman Demirel Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, 19(2), 137–157, 2014.
- [65] N. Usslepp, N. Hübner, G. Stoll, M. Spengler, U. Trautwein, B. Nagengast, “RIASEC interests and the Big Five personality traits matter for life success-But do they already matter for educational track choices?” *Journal of Personality*, 88(5), 1007-1024, 2020.
- [66] G. Stoll, S. Rieger, o. Lüdtke, B. Nagengast, U. Trautwein, B. W. Roberts, “Vocational interests assessed at the end of high school predict life outcomes assessed 10 years later over and above IQ and Big Five personality trait”, *Journal of Personality and Social Psychology*, 113(1), 167-184, 2017.
- [67] P. M. Podsakoff, S. B. MacKenzie, J.-Y. Lee, N. P. Podsakoff, “Common method biases in behavioral research: a critical review of the literature and recommended remedies”, *Journal of Applied Psychology*, 88 (5): 879-903, 2003.