

# Kişiselleştirilmiş Yabancı Dil Öğrenimi İçin Makine Öğrenmesi Yöntemleriyle İlgili Alanı Tahmini

## Prediction of Interest Through Machine Learning Methods for Personalized Foreign Language Learning

Yasin GÖRMEZ, Kübra OKUMUŞ DAĞDELER, Merve KAVUKLU

### ÖZ

Küreselleşen dünyada yabancı dil bilmenin önemi giderek artmaktadır. Dil öğretimi zorluklarını azaltmak için önemli yöntemlerden biri de, teknoloji dünyasındaki gelişmeler ile birlikte daha kolay yönetebilir hâle gelen kişiselleştirilmiş öğrenim yaklaşımlarıdır. Kişiselleştirilmiş öğrenim sayesinde aynı sınıf ortamında bile, her bireyin istek ve ihtiyaçlarına göre yöntem ve materyal sunulabilmektedir. Dil öğretiminde, içeriklerin kişilerin ilgi alanlarına uygun olarak sunulmasının öğrenme verimini artıracığı düşünülmektedir. Bu kapsamda çalışmada, kişiselleştirilmiş İngilizce öğretiminde alt yapı olarak kullanılmak üzere makine öğrenmesi yöntemleri ile bireylerin ilgi alanı tahmini yapılmıştır. Çalışmada, öncelikli olarak bir anket tasarlanarak farklı sektörlerden 164 kişiye uygulanmıştır. Tasarlanan ankette kişilerin birden fazla işaretleme yapabilecekleri, istedikleri kadar seçim yapabilecekleri seçeneklerden oluşan 11 soru ve en sonunda ilgi alanlarını doğrudan işaretleyebilecekleri bir bölüm bulunmaktadır. Birey en az biri zorunlu olmak üzere teknoloji, sağlık, iş yaşamı, farklı kültürler, spor ve güzel sanatlar ilgi alanlarından dilediği kadarını seçebilmektedir. Toplanan bu veriler matematiksel hâle dönüştürülerek k-en yakın komşu, rastgele orman ve yapay sinir ağı yöntemleri ile analizler yapılmıştır. Kullanılan yöntemlerim parametre optimizasyonu için geleneksel ızgara arama yönteminden daha kısa sürede daha iyi sonuçlar üreten Bayesian optimizasyon yönteminden faydalanılmıştır. Bir kullanıcı birden fazla ilgi alanı seçebildiği için tüm makine öğrenmesi modelleri çoklu etiket tahmini yaklaşımı ile oluşturulmuştur. Bu bağlamda her bir kişi için ilgi duyuyor ve duymuyor olacak şekilde 6 ilgi alanı için ayrı ayrı tahmin yapılmış ve başarı oranı da bu durum göze alınarak hesaplanmıştır. Analiz sonuçları incelendiğinde en iyi başarı oranı %78.12 ile rastgele orman algoritması ile elde edilmiştir. Bu sonucun tasarlanacak sistem için yeterli olduğu, veri sayısının artırılması ile birlikte de daha iyi sonuçlar elde edileceği öngörülmektedir.

**Anahtar Sözcükler:** Kişiselleştirilmiş öğrenim, Makine öğrenmesi, Optimizasyon, Yabancı dil eğitimi, İlgi alanı tahmini

Görmez Y., Okumuş Dağdeler K., & Kavuklu M., (2022). Kişiselleştirilmiş yabancı dil öğrenimi için makine öğrenmesi yöntemleriyle ilgi alanı tahmini. *Yükseköğretim ve Bilim Dergisi/Journal of Higher Education and Science*, 12(1), 111-121. <https://doi.org/10.5961/higheredusci.982740>

Yasin GÖRMEZ (✉)

ORCID ID: 0000-0001-8276-2030

Sivas Cumhuriyet Üniversitesi, İktisadi İdari Bilimler Fakültesi, Sivas, Türkiye  
Sivas Cumhuriyet University, Faculty of Economics and Administrative Sciences, Sivas, Turkey  
ysngrmz@gmail.com

Kübra OKUMUŞ DAĞDELER

ORCID ID: 0000-0002-3781-3182

Sivas Cumhuriyet Üniversitesi, Eğitim Fakültesi, İngiliz Dili Eğitimi Bölümü, Sivas, Türkiye  
Sivas Cumhuriyet University, Faculty of Education, Department of English Language Teaching, Sivas, Turkey

Merve KAVUKLU

ORCID ID: 0000-0002-3781-3182

Detay Teknoloji A. Ş. Ar-Ge Merkezi  
Detay Technology Inc. R&D Center

Geliş Tarihi/Received : 14.08.2021

Kabul Tarihi/Accepted : 21.03.2022



Bu eser "Creative Commons Atıf-GayriTicari-4.0 Uluslararası Lisansı" ile lisanslanmıştır.

**ABSTRACT**

In the globalizing world, the importance of knowing a foreign language is increasing. One of the important methods to reduce language teaching difficulties is personalized learning approaches, which have become easier to manage with the developments in the world of technology. Thanks to personalized learning, educational methods and materials can be presented based on the interests and needs of each individual, even in the same classroom environment. It is known that presenting the learning contents in accordance with the interests of the learners will increase the efficiency of learning. Thus, in this study, the individuals' areas of interest were estimated through machine learning methods in order to use it as an infrastructure for personalized foreign language learning. A questionnaire was designed and applied to 164 people from different sectors. In the questionnaire, there were 11 questions consisting of options that people could choose as much as they want, and a section where they could choose their area of interest. Individuals could choose areas of interest as much as they want from technology, health, business life, different cultures, sports and fine arts, at least one of which is compulsory. The collected data were transformed into mathematical form and analyzed with k-nearest neighbor, random forest and artificial neural network methods. The Bayesian optimization method, which produces better results in a shorter time than the traditional grid search method, was used for parameter optimization of methods used. Since a user could select more than one area of interest, all machine learning models were built with a multi-label prediction approach. In this context, separate estimations were made for 6 areas of interest for each person, whether they were interested or not, and the success rate was calculated considering this situation. When the analysis results were examined, the best success rate was obtained with the random forest algorithm with 78.12%. It was foreseen that this result would be sufficient for the system to be designed and better results would be obtained with the increasing number of data.

**Keywords:** Personalized Learning, Machine Learning, Optimization, Foreign Language Education, Interest Prediction

**GİRİŞ**

Bilindiği üzere, ülkemizde yabancı dil eğitimi hem öğrenci hem de öğretmenleri zorlayan bir süreç olma özelliğini hâlâ sürdürmektedir. Yabancı dil eğitiminde yeni yöntem ve stratejiler baz alınarak müfredat ve materyallerde yapılan güncellemeler de henüz isteneni verememiştir. Özellikle de devlet okullarında verilen yabancı dil dersleri öğrencilerin okuma, yazma ve dinleme gibi gerekli becerileri kazanmalarına yardımcı olmamaktadır (Oktay, 2015). English Proficiency Index'ine göre Türkiye, İngilizce yeterliliği konusunda düşük yeterlilik kategorisinde bulunmakta olup 100 ülke arasından 69. sıradadır (*EF EPI 2020 – Turkey*, t.y.). Öte yandan, küreselleşen dünyada İngilizcenin ortak iletişim dili olması İngilizce bilmenin önemini her geçen gün daha da artırmaktadır. Bu nedenle, İngilizce iletişim becerilerinin kazandırılmasında iyileştirmeye ihtiyaç vardır. Her bireyin farklı bilgi, beceri ve ilgi alanları olduğu ve eğitim sürecinde bunları göz önünde bulundurmanın gerekliliği uzun yıllardır bilinmektedir. Öğrencinin kendi ilgi alanı ve bireysel ihtiyaçlarına yönelik dil eğitiminin öğrenci başarısını arttıracağına inanılmaktadır. Bu nedenle, bu çalışmada İngilizce eğitiminde kişiselleştirilmiş öğrenim konusuna odaklanılmıştır.

Basham ve ark. (2016), kişiselleştirilmiş öğrenimin gerçekte ne olduğunun çok az kişi tarafından bilindiğine ve daha da az sayıda insanın kişiselleştirilmiş öğrenimi uygulayabildiğine inanmaktadır. Bulger'e (t.y.) göre, kişiselleştirilmiş öğrenim tanımları, öğrenci merkezli sınıflar, öğrenme yönetim sistemleri, özelleştirilmiş ara yüzler ve uyarlanabilir öğretmenler gibi birçok olasılığı içermektedir. Bu nedenle kişiselleştirilmiş öğrenimi anlamak ve açıklamak çok kolay olmamaktadır. Traxler (2007) kişiselleştirilmiş öğrenimi, öğrenmenin geliştirildiği ve desteklendiği yerlerde çeşitliliği, farklılığı ve bireyselliği tanıyan öğrenme olarak tanımlamaktadır. Eğitimde kişiselleştirme, öğrencilerin farklı yerlerde öğrenmeyi deneyimlemelerini, ken-

di öğrenmelerini gözden geçirmelerini ve beceri ve yetenekleriyle ilgili eğitim kaynaklarına erişmelerini sağlar (Petersen ve ark., 2009). Peng, Ma ve Spector'a (2019) göre, kişiselleştirilmiş öğrenmenin anahtarı bireyselliktir. Kişiselleştirilmiş öğrenmenin temel unsurları ise bireysel farklılıklar, kişisel ihtiyaçlar ve kişisel gelişimdir. Literatürde, kişiselleştirilmiş dil öğreniminde bazı mobil veya web tabanlı sistemler tasarlanmıştır (örn. C.-M. Chen & Chung, 2008; C.-M. Chen & Hsu, 2008; Hsu ve ark., 2013; Petersen ve ark., 2009) ve birkaç derleme makalesi yazılmıştır (C.-M. Chen ve ark., 2006; C.-M. Chen & Li, 2010; Hsieh vd., 2012; Hsu ve ark., 2013; Qoussini & Bt Jusoh, 2014; Shuib ve ark., 2015; Yarandi ve ark., 2011). Örneğin, Ding (2018) İngilizce öğretimi için, öğrenen hakkında farklı bilgiler toplayarak kişiselleştirilmiş elektronik kitaplar oluşturmuştur. Petersen ve ark. (2009) ise, mobil cihazlar aracılığıyla kullanılabilen PALLAS adlı kişiselleştirilmiş ve bağlamsallaştırılmış bir sistem geliştirmiştir. Bu sisteme içerikler öğretmen tarafından yüklenmekteydi. Sistemde kişiselleştirilme için hem çevresel bilgiler hem de kullanıcıların yaş, seviye, ana dili ve ilgi alanı gibi bilgiler kullanılmıştır.

Jung ve Graf (2008), kişiselleştirilmiş kelime oyunları yoluyla web tabanlı bir kelime öğrenimi sistemi önermiştir. Sistem oyun aracılığıyla öğrencilerin mevcut sözcük dağarcığı hakkında bilgi toplamaktaydı. Yine kelime öğretimi için, Chen ve Chung (2008) kelime temelli kişiselleştirilmiş mobil uygulama geliştirirken, Yeung ve Lee (2018) kelime bağlamında en uygun okuma materyalini sunma için metin düzeltme algoritması önermiştir. Jeong ve ark., (2012) ise kelime ve dilbilgisi temelli kişiselleştirilmiş öğrenim ders planlayıcısı sistemi geliştirmiştir. Çalışmada, bu sistemin öğrenme etkililiğini ve öğrenci memnuniyetini artırdığı bulgulanmıştır. Griol ve ark. (2014) multimodal etkileşimli ajan kullanarak, dilbilgisi, kelime ve dinleme etkinlikleri içeren bir kişiselleştirilmiş dil eğitimi sistemi oluşturmuştur. Öte yandan, dilbilgisi üzerine çalışan Troussas ve ark. (2018)

İngilizce ve Fransızca öğretimi için makine öğrenmesi ve bulanık mantık kullanarak kişiselleştirilmiş bir sistem tasarlamıştır. Bu sistem yardımcı fiillerle ilgili hata tespiti içermektedir.

İngilizce okuma becerisini geliştirmek için tasarlanan kişiselleştirilmiş sistemlerin öğrenenlerin okuma becerisini artırdığı görülmüştür. Örneğin; Yao (2017) konum tabanlı bir öneri sistemini genel İngilizce (konuşma, okuma ve kelime) öğretimi için geliştirip deneysel bir çalışma ile test etmiştir. Sonuçlar, sistemin öğrenen ilgisini ve öğrenme etkinliğini artırdığını göstermiştir. Benzer şekilde, Hsu ve ark.'nın (2013) çalışması; kişiselleştirilmiş öneri sistemiyle çalışan deney grubunun okuma becerisi kazanımında kişiselleştirilmiş öneri sistemine sahip olmayan sistemle çalışan öğrencilere göre daha başarılı olduğunu bulgulamıştır. Hsieh ve ark. (2012) öğrenene uygun İngilizce makaleyi önermek için kümülatif öğrenen profilini kullanan kişiselleştirilmiş İngilizce makale önerme sistemi geliştirmiştir. Sistem, bulanık çıkarım mekanizmaları, bellek döngüsü güncellemeleri, öğrenen tercihleri ve analitik hiyerarşi süreci kullanmaktaydı. Öğrenenler ile test edilen sistemin başarılı olduğunu gözlemlenmiştir. Ayrıca, Wu ve ark. (2014) RFID tabanlı konum duyarlı bir teknoloji kullanarak kullanıcının bulunduğu konumla ilgili okuma materyalleri sunan bir sistem tasarlamıştır.

Chen ve ark (2021) yapmış oldukları derleme çalışmasında kişiselleştirilmiş dil öğrenimi konularının, kişiselleştirilmiş dilbilgisi öğrenimi, kişiselleştirilmiş dil öğrenim ağı, dil öğrenimi için kişiselleştirilmiş öneri sistemi, oyun ortamlarında kişiselleştirilmiş dil öğrenimi, dil öğrenimi için kişiselleştirilmiş geri bildirim, kaygı ve kişiselleştirilmiş öğrenim ve mobil destekli kişiselleştirilmiş dil öğrenimi olduğunu tespit etmişlerdir.

Alanyazındaki bu çalışmalar dil öğreniminin farklı alanlarına yönelmiş olup kişiselleştirme için farklı teknikler kullanmıştır. Ancak, dil öğreniminin temel becerilerinden olan ve etkili iletişimin birincil koşullarından olan dinleme becerisi üzerine yoğunlaşan bir çalışma bulunmamaktadır. Ayrıca mevcut çalışmalar, sistem tasarımını baz almış olup, tasarım öncesi makine öğrenmesi gibi bir yöntemle ön tahmin çalışma verisi bulunmamaktadır. Oysa ki, etkili bir kişiselleştirilmiş sistem tasarımı için ön kişisel verilerin toplanması oldukça önemlidir. Ayrıca, bazı çalışmalar konum gibi kişiselleştirmeleri baz alırken bu çalışma öğrenenlerin genel ilgi alanlarını (sağlık, teknoloji, iş yaşamı, güzel sanatlar, farklı kültürler ve spor) temel almıştır.

Ülkemizde de artık farklı öğretim tekniklerine ve zekâ türlerine göre düzenlenmiş hem basılı hem de dijital kaynaklar kullanılmaktadır. Ancak her bireye aynı şekilde ve aynı düzeyde sunulan bu kaynakların kullanıldığı bir eğitim sürecini kişiselleştirilmiş öğrenim olarak adlandırmak mümkün değildir. Piyasada bulunan hem mobil hem de web tabanlı dil öğrenim uygulamaları bireysel ilgi ve ihtiyaçlara yeterince hitap etmemektedir. İngilizce dinleme becerisi geliştirmek için piyasada Ello, Esl-Lab, Oxford, British Council, IELTS listening, All ears English podcast, LearnEnglish Podcasts, Listen English ve İngilizce Dinleme Oynatıcısı gibi birçok web sitesi ve mobil uygulama bulunmaktadır. Ancak bu ürünler kişiselleştirilmiş öğrenmeyi sağlamamaktadır. Farklı konulardan dinleme metinlerini içerse de kullanıcı ilgisini tahmin ederek ona özgü bir içerik getirmemektedir. Bu

nedenle, öğrenciye özgü etkinlik ve konuların yer alacağı, her zaman ve her yerde kullanılabilir mobil ve web tabanlı uygulamalara ihtiyaç vardır. Bireye hitap edecek bu uygulamaların oluşturulmasında gerekli ihtiyaç analizlerini yapmak için daha akılcı bir yöntem ihtiyacı duyulmaktadır.

Kişiselleştirilmiş bir eğitim için mevcut ve gelecek öğrencilerin hepsini bir veri tabanına kaydetmek mümkün değildir. Bu nedenle her bir öğrenci için nasıl bir eğitim materyali kullanılması gerektiği standart bir uygulamayla çözülememektedir. Ancak öğrencileri belirli özelliklerini kullanarak gruplamak ve aynı gruptaki öğrenciler için de aynı eğitim materyalini kullanmak mümkündür. Lineer bir düzlemde çözümü olmayan problemler için sıklıkla kullanılan makine öğrenmesinden yararlanarak gruplama yoluyla kişiselleştirilmiş öğrenimin sağlanması hem mevcut hem de gelecek öğrencilerin dil becerisini geliştireceği düşünülmektedir. Makine öğrenmesi yöntemleri, nesnelere interneti (Stergiou & Psannis, 2017; Xu ve ark., 2019), siber güvenlik (Kim ve ark., 2016; Li & Liu, 2011; Shone ve ark., 2018), biyoenformatik (Wei Yang ve ark., 2011), hastalık teşhisi (Janssens ve ark., 2008; Krupp vd., 2012), görüntü işleme (Gardezi ve ark., 2017; Prabusankaral ve ark., 2015; Zhou & Wang, 2016), metin işleme (Ankit & Saleena, 2018; Catal & Nangir, 2017), trafik akış tahmini (Koesdwady ve ark., 2016; Kumar ve ark., 2013) ve maaş tahmini (Hill & Jolly, 2012; Thacker, 1995) gibi birçok alanda kullanılmış ve bu yöntemler ile başarılı sonuçlar elde edilmiştir. Makine öğrenmesi analizleri ve yapay zekâ yöntemleri baz alınarak oluşturulan teknoloji destekli İngilizce öğreniminin, alanda büyük bir boşluğu dolduracağı ön görülmektedir.

Tüm bu anlatılanlardan dolayı çalışmada, makine öğrenmesi ve yapay zekâ destekli bir sistemin alt yapısını oluşturmak için k-en yakın komşu, yapay sinir ağları, rastgele orman algoritmaları ile makine öğrenmesi analizleri yapılmıştır. Literatürden farklı olarak ilgi alanları tahmini yapan bir yöntem geliştirilmiştir. Elde edilen ilgi alanı tahmin modelinin, bireyin ilgi alanına özgü içerikler sunulmasında kullanılması düşünülmektedir. Bu sayede bireyin kendi ilgi alanına uygun içeriğe daha iyi odaklanarak, İngilizce bilgi ve becerisinin geliştirilmesi hedeflenmektedir. İlgi alanı tahmini için literatürde bir veri seti bulunmamaktadır (bilgimize göre). Bu kapsamda bir anket hazırlanarak veri toplanmıştır. Bu yönü ile de literatürden farklılaşan çalışma, ilgili alanda yapılacak diğer çalışmalar temel olma niteliği de taşımaktadır.

## YÖNTEM

### Veri Seti

Çalışma ile kişilerin ilgi alanlarını tahmin eden bir makine öğrenmesi modeli geliştirilmesi hedeflenmektedir. Bu kapsamda öncelikli olarak 11'i öznitelik sonuncusu sınıf bilgisi olmak üzere 12 sorudan oluşan bir anket hazırlanmıştır. Anket tasarlanma aşamasında Mesleki Yönelim Envanteri'nden yararlanılmıştır (Kepçeoğlu, 2019). Tasarlanan anket farklı şehir ve sektörden 164 kişiye uygulanmıştır. Anket için sorular, her sorunun seçenekleri ve seçeneklerin tercih edilme oranları Tablo 1'de sunulmuştur. Her bir soru için birden fazla seçeneği seçmek soruda da belirtildiği gibi mümkün olmaktadır.

**Tablo 1:** Veri Seti Soruları, Seçenek ve Seçeneklerin Tercih Edilme Oranları

Sorular	Seçenek	Tercih Edilme Oranı
Cinsiyet	Erkek Kız	%58.5 %41.5
Kendinize uygun olduğunu düşündüğünüz en az bir seçeneği işaretleyiniz	Maliyet hesapları üzerine kursa katılmak İlk yardım kursuna katılmak Bir spor kursuna katılmak Farklı mutfak kültürleri kursuna katılmak Tiyatro kursuna katılmak Bilgisayar kursuna katılmak Resim kursuna katılmak Robotik kodlama kursuna katılmak Enstrüman kurslarına katılmak	%8.5 %25.6 %36.6 %29.3 %22.6 %30.5 %22 %17.1 %56.1
Kendinize uygun olduğunu düşündüğünüz en az bir seçeneği işaretleyiniz	Bir şehrin sağlık kurumlarını gezmek Bir şehrin tiyatro ve sinema salonlarını gezmek Bir şehrin stadyumlarını gezmek Bir şehrin müzelerini gezmek Bir şehrin idari binalarını gezmek Şehirdeki teknoloji marketlerini gezmek Şehirdeki fitness merkezlerini gezmek Şehrin konser salonlarını gezmek	%9.1 %50 %15.2 %68.9 %7.9 %35.4 %12.8 %39.6
Kendinize uygun olduğunu düşündüğünüz en az bir seçeneği işaretleyiniz	Güncel teknolojilerin tartışıldığı bir gruba katılmak Müzik eserlerinin analiz edildiği bir gruba katılmak Finans konularının tartışıldığı bir gruba katılma Spor müsabakalarının tartışıldığı bir gruba katılmak Asya ya da Avrupa kültürlerin tartışıldığı bir gruba katılmak Kanser tedavisindeki gelişmelerin tartışıldığı bir gruba katılmak	%39 %37.2 %14 %17.7 %50 %24.4
Kendinize uygun olduğunu düşündüğünüz en az bir seçeneği işaretleyiniz	Spor müsabakası izlemek Film izlemek İş hayatı üzerine bir program izlemek Gezi programı izlemek Müzik programı izlemek Teknoloji ile ilgili bir program izlemek	%24.4 %74.4 %23.2 %51.8 %45.1 %39
Kendinize uygun olduğunu düşündüğünüz en az bir seçeneği işaretleyiniz	Teknolojik gelişmelerle ilgili makale yazmak Sanatsal çalışmalarla ilgili makale yazmak Farklı kültürler hakkında makale yazmak Tıp alanında makale yazmak İş hayatı hakkında makale yazmak Sportif faaliyetler hakkında makale yazmak	%32.9 %37.8 %56.7 %3 %22.6 %18.9
Kendinize uygun olduğunu düşündüğünüz en az bir seçeneği işaretleyiniz	Ünlü sporcuların hayat hikâyelerini okumak Ünlü doktorların hayat hikâyelerini okumak Ünlü iş adamlarının hayat hikâyelerini okumak Ünlü bilimcilerin hayat hikâyelerini okumak Ünlü sanatçıların hayat hikâyelerini okumak Farklı ülkelerde yaşayan insanların hayat hikâyelerini okumak	%20.7 %17.1 %29.9 %30.5 %47 %66.5
Kendinize uygun olduğunu düşündüğünüz en az bir seçeneği işaretleyiniz	Edebi eserlerin tartışıldığı bir gruba katılmak Amerika kültürlerinin tartışıldığı bir gruba katılmak Avusturalya kültürlerin tartışıldığı bir gruba katılmak Yönetim konusunun tartışıldığı bir gruba katılmak Tıbbi araştırmaların tartışıldığı bir gruba katılmak Yapay zekânın tartışıldığı bir gruba katılmak. İletişim teknolojilerinin tartışıldığı bir gruba katılmak Ünlü ressamın eserlerinin tartışıldığı bir gruba katılmak	%34.8 %32.3 %24.4 %20.7 %12.2 %46 %40.9 %25

Tablo 1: Devam

Sorular	Seçenek	Tercih Edilme Oranı
Kendinize uygun olduğunu düşündüğünüz en az bir seçeneği işaretleyiniz	Bir üniversitede kütüphaneyi gezmek	%56.7
	Üniversite işlem merkezini görmek	%20.7
	Rektörlük ve idare binalarını dolaşmak	%13.4
	Tiyatro ve konser salonlarını incelemek	%50
	Üniversite öğrenci sağlık merkezini görmek	%15.2
	Üniversitenin basketbol, futbol vs. sahalarını görmek	%34.1
	Üniversitedeki uluslararası ilişkiler ofisini görmek	%30.5
	Üniversitedeki fitness merkezini görmek	%24.4
Kendinize uygun olduğunu düşündüğünüz en az bir seçeneği işaretleyiniz	Güzel sanatlar alanından biri ile tanışmak	%48.8
	Spor alanından biri ile tanışmak	%29.9
	Tıp ve sağlık alanından biri ile tanışmak	%26.8
	Edebiyat ve dil alanından biri ile tanışmak	%45.7
	Teknoloji alanından biri ile tanışmak	%43.3
	İdare ve ekonomi alanından biri ile tanışmak	%20.1
	Farklı uyruğa sahip insanlarla tanışmak	%64.6
Kendinize uygun olduğunu düşündüğünüz en az bir seçeneği işaretleyiniz	İleride, güzel sanatlara yönelmek	%36.6
	Yabancı dil alanına yönelmek	%70.1
	İdare ve ekonomi alanına yönelmek	%20.1
	Tıp ve sağlık bilimleri alanına yönelmek	%6.7
	Teknik bilimler alanına yönelmek	%19.5
	Spor alanına yönelmek	%29.9
İlgi alan(lar)ınızı işaretleyiniz.	Teknoloji	%50
	Sağlık	%24.4
	İş yaşamı	%30.5
	Farklı kültürler	%62.8
	Spor	%47
	Güzel sanatlar	%56.7

Daha sonra toplanan bu verinin rastgele %20'u test %10'u validasyon ve kalan %70'i validasyon için eğitim olmak üzere parçalara ayrılmıştır. Her bir soruya birden fazla cevap verilmesi mümkün olmasından dolayı, bir kişi tarafından verilen cevapların matematiksel gösterimini yapmada ikilik taban sistemi kullanılmıştır. Bu kapsamda, soruda yer alan seçenek sayısı kadar basamağa sahip bir sayı üretilmiştir. Her bir basamak soruda yer alan bir seçeneği temsil etmektedir. Her örnek için öznitelikler, ilgili soruda seçili olan seçeneği temsil eden basamak eğer o seçenek seçili ise 1 değilse 0 olacak şekilde matematiksel gösterime çevrilmiştir. Örneğin 5 seçeneğe sahip bir soru için, 1. ve 3. seçenekler tercih edilmiş ise, o örnek için sorunun matematiksel gösterimi 10100 olmaktadır.

### Sınıflama

Önceki bölümlerde de anlatıldığı gibi, bireylerin ilgi alanlarının tahmin edilmesi hedeflenmektedir. Veri toplama aşamasında bir kişi birden fazla ilgi alanı seçebilmektedir. Bu nedenle çalışmada 3 farklı sınıflama yöntemi çoklu etiket sınıflama tekniğinden faydalanılarak kullanılacaktır.

### K-en Yakın Komşu

K-en yakın komşu algoritması, girdi olarak verilen bir veriyi, sistemde bulunan veri kümesinde yer alan en yakın gruba atayan bir sınıflama algoritmasıdır (Peterson, 2009). Bu amaçla önce-

likli olarak verilen bir girdinin, eğitim kümesinde yer alan tüm verilere olan uzaklığı hesaplanmaktadır. Bu aşamada Öklid, Manhattan ve Minkowski gibi uzaklık formülleri kullanılmaktadır. Daha sonra eğitim kümesinde yer alan ve verilen girdiye en az uzaklıktaki  $k$  kadar eleman kullanılarak bir alt küme oluşturulmaktadır. Son aşamada ise girdi, bu alt kümede yer alan en fazla sınıfa atanarak algoritma sonlandırılmaktadır.

Diğer yöntemlerden farklı olarak, tahmin aşamasında kullanmak üzere, eğitim verisini saklamaktadır. Bu nedenle algoritmada tahmin aşamasında diğer yöntemlere göre daha fazla kaynak gerektirmektedir. Bunun yanı sıra makine öğrenmesi algoritmaları için eğitim veri setindeki örnek sayısının fazla olmasının, model başarı oranını etkileyen en önemli etkenlerden biri olduğu bilinmektedir. K-en yakın komşu algoritması için her tahmin aşamasında bahsedilen hesaplamaların yapılması ise örnek sayısı arttıkça modelin yavaşlamasına neden olmaktadır. Veri sayısında çok fazla artış meydana geldiğinde belirtilen yavaşlığın dramatik şekilde artıyor olması, yöntemin fazla veri sayısı ile kullanılamamasına neden olmaktadır. Ancak küçük veri setleri için çok iyi başarı oranlarını kısa sürede elde edebilmesinden dolayı, literatürde sıklıkla kullanılmaktadır.

### Rastgele Orman

Rastgele orman algoritması, topluluk öğrenmesi konseptini

kullanan gözetimli öğrenme yöntemlerinden biridir (Bonissone ve ark., 2010). Yöntemde çeşitli karar ağaçları eğitilmekte ve bu ağaçlar ile elde edilen sonuçların ortalaması alınarak nihai sonuç elde edilmektedir. Modelin başarı oranının daha yüksek olması için, her ağaçtan gelen sonuçlar arasında korelasyonun az olması gerekmektedir. Diğer algoritmalara göre daha az eğitim zamanı alması, büyük ya da küçük veri seti fark etmeksizin daha verimli sonuçlar elde etmesi ve verilerin büyük çoğunluğu kayıp olduğunda bile iyi sonuçlar elde edebiliyor olması algoritmanın en büyük avantajları olarak düşünülebilir. Karar ağaçlarının çalışma adımları aşağıdaki aşamalarla açıklanabilir:

1. Eğitim setinden rastgele N veri seçilir.
2. Seçilen veriler kullanılarak bir karar ağacı oluşturulur.
3. Aşama 1 ve 2 kullanılarak M adet karar ağacı oluşturulur.
4. Yeni verilerin sınıfları, ilk 3 aşamada elde edilen karar ağacı tahminlerinin çoğunluk oylaması (majority voting) ile bulunur.

### Yapay Sinir Ağları

Yapay sinir ağları (YSA), insan sinir sistemini taklit ederek geliştirilen gözetimli öğrenme yöntemlerinden biridir. Basit bir sinir ağı, girdi, gizli ve çıkış olmak üzere en az 3 katmandan oluşmaktadır (Hopfield, 1988). Her bir katmanda nöron adı verilen yapılar bulunmaktadır. Girdi katmanı verilerin ilk okunduğu katmandır ve veri setindeki öznitelik sayısı kadar nöron içermektedir. Gizli katman sayısı ve her bir gizli katmanda bulunan nöron sayısı değişiklik gösterebilmekte olup, başarı oranını etkileyen en önemli faktörlerden biridir. Bu nedenle nöron sayısı değeri birçok çalışmada optimize edilmektedir. Çıkış katmanı ise tek nöron içerebildiği gibi, sınıf sayısı kadar nöronda içerebilmektedir, ancak her iki durumda aktivasyon fonksiyonu uygun olarak belirlenmelidir. Tam bağlı bir yapay sinir ağı modelinde, bir katmanda yer alan her bir nöron takip eden katmanda yer alan tüm nöronlara bağlıdır. Bu bağı sağlayan yapılara ise ağırlık adı verilmektedir ve model ile asıl öğrenilmek istenen bu ağırlık değerleridir. İleri besleme ve geri yayılım olmak üzere iki ana aşaması bulunmaktadır. İleri besleme kısmında, nörondaki girdi değeri ile o nöronu takip eden katmandaki nörona bağlanan ağırlık değeri çarpılır. Daha sonra elde edilen değer bir aktivasyon fonksiyonundan geçirilerek takip eden katmandaki nöronun değeri hesaplanır. Bu işlem çıkış katmanına kadar yapılarak nihai değer hesaplanır ve bu değere göre sınıf belirlenir. Geri yayılım aşamasında ise tahmin hata durumuna göre modelde yer alan ağırlıklar güncellenir. İleri besleme ve geri yayılım algoritmasının belirli sayıda tekrar edilmesi ile eğitim sonlandırılır.

### Bayesian Optimizasyon

Makine öğrenmesi yöntemleri özgün parametrelere sahiptir ve bu parametrelerin doğru kullanımı, başarı oranını etkileyen önemli faktörlerden biridir. Parametre optimizasyonu yapmak makine öğrenmesi modellerinin başarı oranları için hayati öneme sahip olsa da, parametre sayısına bağlı olarak optimizasyon, çok fazla zaman alabilmektedir. Özellikle de parametre çeşidinin fazla ve parametre uzayının geniş olmasından ötürü yapay sinir ağı modellerinde bu durum daha büyük sorun

hâline gelmektedir (Görmez ve ark., 2021). Optimizasyon işleminde ızgara arama (grid search) yöntemi de kullanılmaktadır. Bu yöntemde her bir parametre için değer listesi oluşturulmakta ve her bir değer kullanılarak eğitim işlemi yapılmaktadır. Daha sonra modelin başarı oranı validasyon seti kullanılarak ölçülmektedir. Son aşamada en yüksek başarı oranını veren parametre kullanılarak model eğitilmekte ve test veri seti kullanılarak başarı oranı hesaplanmaktadır. Bu yöntemin ise iki ana problemi bulunmaktadır. İlk olarak parametre çeşidinin fazla olması durumunda eğer parametre uzayı da geniş ise iç içe döngülerden dolayı optimizasyon zamanı büyük ölçüde artmakta hatta bazen hesaplanamaz duruma gelmektedir. Optimizasyon zamanını azaltmak için parametre uzayının dar tutulması durumunda ise yeterli seviyede optimizasyon yapılamamaktadır. Diğer bir problem ise, verilen listede yer almayan değerler kullanılmadığı için hassasiyet azalmaktadır. Örneğin yapay sinir ağı parametresi olan öğrenme oranı (Learning Rate - lr) için oluşturulan liste {0.1, 0.3, 0.01, 0.03} olsun. Bu durumda arada kalan 0.15, 0.2342 gibi birçok değer modelde denenememekte bu da hassasiyeti azaltmaktadır. Bu sorunlara çözüm olarak çalışmada Bayesian Optimizasyon tekniğinden faydalanılmıştır. Bu yöntemde parametre uzayı listesi vermek yerine, maksimum ve minimum değerleri verilmektedir. İlk başta belirtilen aralıktan bir değer seçilerek modelin başarı oranı ölçülmektedir. Daha sonra, yeni denemeler yoluyla en uygun parametre değerlerinin bulunması çok değişkenli Gauss dağılımından faydalanılarak yapılmaktadır. Bir Gauss çekirdeği, parametreler arasındaki fonksiyon değerlerinin kovaryansını belirtir. Her yinelemede, bir parametrenin sonraki değeri, Gauss önceliği üzerinden edinme işlevi aracılığıyla seçilir.

### BULGULAR

Analiz aşamasında öncelikli olarak etiket bilgileri ikili sistemde temsil edilecek şekilde düzenlenmiştir. Her bir örnek için  $6 \times 1$  uzunluğunda bir vektör oluşturulmuştur. 6 değeri sınıf sayısını belirtmekte olup vektörün her bir elemanı, veri setindeki bir sınıfı temsil etmektedir. Bu nedenle ilgili örnekte seçilmiş olan ilgi alanları için vektör değeri 1, diğerleri için 0 olarak ayarlanmıştır. Sonuç olarak  $M \times 6$  ( $M$  örnek sayısını temsil etmektedir) uzunluğunda bir etiket matrisi oluşturulmuştur. Bu vektörde yer alan 1 ve 0 değerleri kişinin seçmiş olduğu ilgi alanlarına göre değişiklik gösterebilmektedir. Böylece çoklu etiket yöntemi ile sınıflama yapılabilmektedir. Veri seti hazırlama bölümünde de bahsedildiği üzere, 32 örnek test verisi, 16 örnek validasyon verisi ve kalan 116 veri validasyon için eğitim verisi olarak ayrılmıştır. Sonuç olarak,  $32 \times 11$ ,  $16 \times 11$  ve  $116 \times 11$  olmak üzere 3 öznitelik matrisi ve  $32 \times 6$ ,  $16 \times 6$  ve  $116 \times 6$  olmak üzere 3 etiket matrisi oluşturulmuştur. Veri setleri oluşturma ve bölme işlemleri yapıldıktan sonra, k-en yakın komşu modelinde  $k$  parametresi, rastgele orman modelinde  $max\_depth$  ve  $n\_estimators$  parametreleri, yapay sinir ağı modelinde ise *öğrenme oranı*, *epoch*, *ara katmandaki nöron sayısı* ( $n\_unit$ ), *dropout oranı* Bayesian optimizasyon yöntemi ile optimize edilmiştir. k-en yakın komşu algoritmasında hesaplama kolaylığı nedeni ile Öklid uzaklığı kullanılmıştır. Bu aşamada k-en yakın komşu ve rastgele orman modelleri için scikit-learn (*scikit-learn: machine learning in Python — scikit-learn 0.24.2 documentation*, t.y.),

**Tablo 2:** Parametre Aralıkları ve Optimum Değerler

Model	Parametre	En Düşük Değer	En Yüksek Değer	Optimum Değer
K-en yakın komşu	k	1	25	12
Rastgele orman	max_depth	1	12	3
Rastgele orman	n_estimators	10	500	147
Yapay sinir ağı	öğrenme oranı	10 <sup>-6</sup>	10 <sup>-1</sup>	0.02937875
Yapay sinir ağı	epoch	5	400	232
Yapay sinir ağı	n_unit	10	250	108
Yapay sinir ağı	dropout oranı	0	0.9	0.2

**Tablo 3:** Optimum Parametreleri Kullanan Modellerin TP, TN, FP ve FN Sayıları

Model	TP	TN	FP	FN
k-en yakın komşu	46	96	21	29
Rastgele orman	49	101	16	26
Yapay sinir ağı	31	104	13	44

**Tablo 4:** TP, TN, FP ve FN Belirlenme Gösterimi

	Bireyin Seçimi	Modelin tahmini
TP	İlgi Duyuyor	İlgi Duyuyor
TN	İlgi Duymuyor	İlgi Duymuyor
FP	İlgi Duymuyor	İlgi Duyuyor
FN	İlgi Duyuyor	İlgi Duymuyor

yapay sinir ağı için keras (*Keras: the Python deep learning API*, t.y.) ve Bayesian optimizasyon için skopt (*skopt module*, 2019) kütüphaneleri python dilinde kullanılmıştır. Modeller, her adımda seçilen parametreler ile validasyon için eğitim verisi kullanılarak eğitilmiş ve validasyon verisinde en yüksek başarı oranını veren parametreler not edilmiştir. Skopt kütüphanesinde *gp\_minimize* fonksiyonu, parametreleri *cq\_func='EI'* ve *n\_calls=100* olacak şekilde ayarlanarak kullanılmıştır. Yapılan araştırmalara göre Bayesian optimizasyon yöntemi EI fonksiyonu ile genellikle ilk 100 çağrıda optimum parametreleri bulmaktadır (Görmez ve ark., 2021). Tablo 2’de model parametrelerinin maksimum, minimum ve optimum değerleri gösterilmiştir.

Tablo 2 incelendiğinde tüm parametrelerde, optimum değerlerin en büyük ya da en küçük değerlere yaklaşmadığı gözlenmektedir. Optimum değerlerin, en düşük ya da en yüksek değerle aynı ya da bu değerlere çok yakın olması durumunda parametreye uzayın genişletilerek yeniden analiz yapılması planlanmıştır. Ancak optimum değerler en yüksek ve en düşük değerlerin arasında yer aldığı için, optimizasyon işleminin yeterli olduğu kanaatine varılmıştır. Parametre optimizasyonu aşamasından sonra validasyon için eğitim ve validasyon veri setleri birleştirilerek 132 x 6 uzunluğunda eğitim veri seti oluşturulmuştur. Daha sonra optimum parametreler kullanılarak model eğitimi yapılmış ve 32 x 11 uzunluğundaki test verisi tahminleri elde edilmiştir. Tüm bu analizlerde yapay sinir ağları ara katmanlarında aktivasyon fonksiyonu *relu*, son katmanda aktivasyon fonksiyonu *sigmoid*, kayıp fonksiyonu (loss function) ortalama

kareler hatası, hata optimizasyon yöntemi *Adam*, *beta1* parametresi 0.95 ve *beta2* parametresi 0.99 olarak ayarlanmıştır. Yapay Sinir Ağı modellerinde ilgili parametrelerin genelde bu değerlere yakın olarak elde edildiği bu nedenle optimizasyona gerek olmadığı kanaatine varılmıştır (Goodfellow ve ark., 2016). Diğer tüm parametreler varsayılan olarak bırakılmıştır. Eğitilmiş modellere ait doğru pozitif (true positive - TP), doğru negatif (true negative - TN), yanlış pozitif (false positive - FP) ve yanlış negatif (false negative - FN) sayıları Tablo 3’te sunulmuştur.

Test verisi örnek sayısı 32 olmasına rağmen, tablo sonuçlarına göre her modelde TP, TN, FP ve FN toplamı 192 olarak görülmektedir. Bunun nedeni, oluşturulan modellerin çoklu etiket yaklaşımıyla tahmin edilmiş olmasıdır. Örneklerin her bir ilgi alanına ilgi duyup duymadığı ayrı ayrı tahmin edilmiştir. Bu nedenle 1 örnekte 6 sınıf değerinin de doğru tahmin edilmesi gerekmektedir. Bu nedenle tahmin edilmesi gereken toplam veri sayısı 32 x 6 yani 192 olmaktadır. TP, TN, FP ve FN değerlerinin nasıl belirlendiği Tablo 4 ile gösterilmiştir. Tablo 3 de elde edilen değerlere bakıldığında TP tahmininde en iyi yöntemin rastgele orman algoritması, TN tahmininde ise yapay sinir ağı algoritması olduğu görülmektedir. İki yöntem arasında karşılaştırma yapıldığında yapay sinir ağı modelinin TN tahmininde çok iyi olmasına rağmen TP tahmininde çok düşük sonuçlar elde ettiği görülmektedir. Buna karşın rastgele orman algoritmasının TN tahmininde yapay sinir ağına çok yaklaştığı görülmektedir.

Modellerin performanslarını daha rahat bir şekilde gözlemleyebilmek için, elde edilen bu değerler kullanılarak başarı oranı (accuracy), kesinlik (precision), duyarlılık (recall) ve doğru negatif oranı skorları hesaplanmıştır (*Precision and recall*, 2017). Hesaplanan bu değerler Tablo 5’te sunulmuştur.

Sonuçlar incelendiğinde doğru negatif oranı hariç diğer tüm alanlarda rastgele orman modelinin en iyi sonucu verdiği gözlemlenmektedir. Bu modelin doğru negatif oranı skoru da, ilgili skorda en iyi sonucu veren yapay sinir ağına göre göz ardı edilebilir seviyedir. Yapay sinir ağı modelinin ise doğru negatif

**Tablo 5:** Optimum Parametreleri Kullanan Modeller İçin Performans Ölçütleri

Model	Başarı Oranı	Kesinlik	Duyarlılık	Doğru negatif oranı
k-en yakın komşu	%73.95	%68.65	%61.33	%82.05
Rastgele orman	%78.12	%75.38	%65.33	%86.32
Yapay sinir ağı	%70.31	%70.45	%41.33	%88.80

oranı hariç, diğer yöntemlere göre düşük sonuçlar elde ettiği gözlemlenmektedir. Özellikle duyarlılık skoru için bu fark çok ciddi seviyelere ulaşmıştır. Bunun nedeninin, veri sayısının az olmasından ötürü öğrenmenin tam olarak yapılamaması olduğu düşünülmektedir. Yapısı gereği yapay sinir ağı modelleri çok fazla sayıda ağırlık katsayısı öğrenmelidir ve bu nedenle fazla veri sayısına ihtiyaç duymaktadır. Veri sayısının artması durumunda yapay sinir ağı modelinin diğer modellere yakın sonuçlar alabileceği hatta geçebileceği düşünülmektedir. Rastgele orman algoritması ile alınan sonuçların, tasarlanması planlanan teknoloji destekli kişiselleştirilmiş İngilizce öğrenimi platformunda kullanılmak üzere yeterli olduğu düşünülmektedir. Veri setinin artırılması durumunda bu sonuçların iyileşebileceği de göz önüne alındığında, makine öğrenmesi yöntemlerinin tasarlanacak platforma büyük katkı sağlayacağı ön görülmektedir.

### SONUÇLAR

Çalışmada kişiselleştirilmiş dil öğreniminde hem mobil hem de web tabanlı olarak tasarlanması planlanan sistem için makine öğrenmesi modellerinin analizleri yapılmıştır. Bu kapsamda öncelikle bir anket tasarlanmış ve farklı sektörlerden 164 bireye uygulanmıştır. Toplanan veriler k-en yakın komşu, rastgele orman ve yapay sinir ağları yöntemleri ile analiz edilmiştir. Parametre optimizasyonu için geleneksel izgara arama tekniğinden daha iyi sonuçları daha kısa sürede elde edebilen Bayesian optimizasyon tekniğinden faydalanılmıştır. Kişilere ilgi alanına uygun içerikler sunulduğunda, dil öğrenme süreçlerine daha iyi odaklanacakları ve böylece öğrenme etkililiğinin artırılabileceği ön görülmektedir. Bu bağlamda, kişilerin ilgi alanlarının en az hata ile tahmin edilmesi, tasarlanacak olan platformda hayati öneme sahiptir. Analiz sonuçları incelendiğinde en iyi sonucu rastgele orman yöntemi ile elde edildiği ve bu sonucun tasarlanması planlanan platform için yeterli olduğu kanısına varılmıştır. Rastgele orman yöntemi ile bulunan % 78.12 değeri uygulanacak çalışma için güvenilir bir sonuç olarak değerlendirilebilir. Çünkü anketin de dahil olduğu birçok veri toplama aracı için kabul edilebilir güvenilirlik katsayısı 70 -.90 aralığıdır (McMillan & Schumacher, 2014). Ancak, ilgi alanı tahmininde başarı oranının %100'e yaklaşmasının platform performansını artıracakları ön görülmektedir. Bu nedenle ilerleyen çalışmalarda daha fazla veri toplanarak analizlerin tekrar edilmesi planlanmaktadır. Yeteri kadar veri elde edilmesi durumunda özellikle yapay sinir ağı modelinin daha iyi sonuçlar elde edeceği, bunun yanı sıra derin öğrenme modellerinin kullanılabilirliği düşünülmektedir. Bunun yanı sıra, tasarlanacak olan platformda kişinin ilgi alanını tahmin etme amaçlı, veri toplama aşamasında kullanılan anket sorularının, platformu kullanacak kişilere tek sefere özgü olarak yapılması planlanmaktadır. Ankette daha az soru sorulması, platformun kullanılabilirliğini artıracaktır. Bu neden-

le sonraki çalışmalarda öznel seçim yöntemleri uygulanarak en az sayıda soru ile en iyi başarı oranı elde eden yöntemler belirlenecektir.

### Extended Abstract

As it is known, foreign language education still continues to be a challenging process for both students and teachers in Turkey. The updates made in the curricula and materials based on new methods and strategies in foreign language education have not yet produced what is desired. Foreign language courses given in especially public schools do not help students acquire necessary skills such as reading, writing and listening (Oktay, 2015). The fact that English is the common language of communication in the globalizing world increases the importance of knowing English. Therefore, there is a need for improvement in the acquisition of English communication skills. It has been known for many years that each individual has different knowledge, skills and interests and it is necessary to consider them in learning process.

It is believed that the language education based on the students' interests and needs will increase student success. Therefore, this study focuses on personalized learning in English education. In this context, machine learning analyzes were made with k-nearest neighbor, artificial neural networks, random forest algorithms to create infrastructure for a machine learning and artificial intelligence supported language learning system. A method that estimated areas of interest was developed, which was different from what was seen in the literature. It was planned that the obtained interest prediction model would be used in presenting content specific to the individual's interest. In this way, it was aimed to improve English skills by focusing on the interests of learners.

There is no data set in the literature (to our knowledge) for area of interest estimation. Thus, a questionnaire was prepared and data were collected. This study, which differs from the literature in this aspect, also carries the feature of being the basis for other studies to be done in the related field.

This study aimed at developing a machine learning model that predicted people's interests. In order to achieve this goal, firstly a questionnaire consisting of 12 questions, 11 of which were attribute information and the last one was class information, was prepared. As developing questionnaire, The Vocational Orientation Inventory (Kepçeoğlu, 2019) was benefited. The questionnaire was applied to 164 people from different cities and sectors. Later, this collected data were divided into parts: 20% for testing, 10% for validation, and the remaining 70% for training for validation. Since it was possible to give more than one answer to each question, the binary base system was used



to represent the answers given by one person. In this context, a number with as many digits as the number of options in the question was generated. Each digit represented an option in the question. Attributes for each sample were converted to mathematical representation with the digit representing the option selected in the relevant question 1 if that option is selected, if not 0.

During the data collection phase, a person could choose more than one area of interest. For this reason, 3 different classification methods were used in the study by using the multi-label classification technique. K-nearest neighbor algorithm is a classification algorithm that assigns an input data to the closest group in the data set in the system. For this purpose, the distance of a given input from all the data in the training set was calculated. At this stage, distance formulas such as Euclid, Manhattan and Minkowski were used. Then, a subset was created by using k elements in the training set and at the least distance from the given input. In the last stage, the algorithm was terminated by assigning the input to the highest number of classes in this subset.

The random forest algorithm is one of the supervised learning methods that uses the concept of community learning. In the method, various decision trees were trained and the final result was obtained by averaging the results obtained with these trees. In order for the model to have a higher success rate, there should be less correlation between the results from each tree. It can be considered as the biggest advantages of the algorithm that it takes less training time than other algorithms, obtains more efficient results regardless of large or small data set, and can obtain good results even when the majority of data is lost.

Artificial neural networks (ANNs) are one of the supervised learning methods developed by imitating the human nervous system. A simple neural network consists of at least 3 layers: input, hidden and output. Each layer contains structures called neurons. The input layer is the layer where the data is read first and contains as many neurons as the number of features in the data set. The number of hidden layers and the number of neurons in each hidden layer can vary, and this is one of the most important factors affecting the success rate. Therefore, the neuron count value is optimized in many studies. The output layer, on the other hand, can contain a single neuron or as many neurons as the number of classes, but in both cases, the activation function should be determined appropriately.

In the phase of analysis, the label information was primarily arranged to be represented in a binary system. A vector of length 6 x 1 was created for each sample. A value of 6 indicates the number of classes, and each element of the vector represents a class in the data set. Therefore, the vector value was set to 1 for the areas of interest selected in the relevant example, and 0 for the others. As a result, a label matrix of length M x 6 (M represents the number of samples) was created. The values of 1 and 0 in this vector may vary as to the areas of interest that the person has chosen. Thus, classification was made through the multi-label method. As mentioned in the data set preparation section, 32 sample test data, 16 sample

validation data and the remaining 116 data were reserved as training data for validation. As a result, 3 feature matrices 32 x 11, 16 x 11 and 116 x 11 and 3 tag matrices 32 x 6, 16 x 6 and 116 x 6 were created. After forming and splitting data sets, k parameter in k-nearest neighbor model, max\_depth and n\_estimators parameters in random forest model, learning rate, epoch, number of neurons in the middle layer (n\_unit), dropout rate in artificial neural network model were optimized through Bayesian optimization method. After the parameter optimization phase, training and validation datasets were combined for validation and a training dataset of 132 x 6 lengths was created. Then, model training was done by using optimum parameters and 32 x 11 test data predictions were obtained. In all of these analyses, the activation function was set to relu in the artificial neural network interlayers, the activation function was sigmoid in the last layer, the loss function was set to mean squares error, the error optimization method was Adam, beta1 parameter was 0.95 and beta2 parameter was 0.99. All other parameters were left by default. As a result of analysis, 73.95% success rate was obtained with the k-nearest neighbor algorithm, 78.12% with the random forest algorithm and 70.31% with the artificial neural network.

It is assumed that when people are presented the content which is suitable for their interests, they will focus better on their language learning and so the efficiency of learning will be affected positively. Thus, predicting people's interests with the least error is important for the platform to be designed. When the results of analysis were examined, it was concluded that the best result was obtained with the random forest method and this result was sufficient for the planned platform. However, it is predicted that the success rate of interest prediction near 100% will increase the platform performance. For this reason, it is planned to collect more data and repeat the analyzes in future studies. It is thought that if enough data is obtained, especially the artificial neural network model will achieve better results, as well as deep learning models can be used. In addition, it is planned that the survey questions used in the data collection phase to predict the interests of people in the platform to be designed will be made only once for the people who will use the platform. Asking fewer questions in the survey will increase the usability of the platform. For this reason, in future studies, feature selection methods will be applied and the methods that have the best success rate with the least number of questions will be determined.

#### KAYNAKLAR

- Ankit, & Saleena, N. (2018). An Ensemble Classification System for Twitter Sentiment Analysis. *Procedia Computer Science*, 132, 937-946. <https://doi.org/10.1016/j.procs.2018.05.109>
- Basham, J. D., Hall, T. E., Carter, R. A., & Stahl, W. M. (2016). An Operationalized Understanding of Personalized Learning. *Journal of Special Education Technology*, 31(3), 126-136. <https://doi.org/10.1177/01626434166660835>
- Bonissone, P., Cadenas, J. M., Carmen Garrido, M., & Andrés Díaz-Valladares, R. (2010). A fuzzy random forest. *International Journal of Approximate Reasoning*, 51(7), 729-747. <https://doi.org/10.1016/j.ijar.2010.02.003>

- Bulger, M. (t.y.). *Personalized Learning: The Conversations We're Not Having*. 29.
- Catal, C., & Nangir, M. (2017). A sentiment classification model based on multiple classifiers. *Applied Soft Computing*, 50, 135-141. <https://doi.org/10.1016/j.asoc.2016.11.022>
- Chen, C.-M., & Chung, C.-J. (2008). Personalized mobile English vocabulary learning system based on item response theory and learning memory cycle. *Computers & Education*, 51(2), 624-645. <https://doi.org/10.1016/j.compedu.2007.06.011>
- Chen, C.-M., & Hsu, S.-H. (2008). Personalized Intelligent Mobile Learning System for Supporting Effective English Learning. *Educational Technology & Society*, 11(3), 153-180.
- Chen, C.-M., Hsu, S.-H., Li, Y.-L., & Peng, C.-J. (2006). Personalized Intelligent M-learning System for Supporting Effective English Learning. *2006 IEEE International Conference on Systems, Man and Cybernetics*, 6, 4898-4903. <https://doi.org/10.1109/ICSMC.2006.385081>
- Chen, C.-M., & Li, Y.-L. (2010). Personalised context-aware ubiquitous learning system for supporting effective English vocabulary learning. *Interactive Learning Environments*, 18(4), 341-364. <https://doi.org/10.1080/10494820802602329>
- Chen, X., Zou, D., Xie, H., & Cheng, G. (2021). Twenty Years of Personalized Language Learning: Topic Modeling and Knowledge Mapping. *Educational Technology & Society*, 24(1), 205-222.
- Ding, L. (2018). Exploration of Key Technologies in a Personalized English Learning System. *International Journal of Emerging Technologies in Learning (IJET)*, 13(07), 85-96. <https://doi.org/10.3991/ijet.v13i07.8789>
- EF EPI 2020 – Turkey. (t.y.). Geliş tarihi 13 Temmuz 2021, gönderen <https://www.ef.com/ca/epi/regions/europe/turkey/>
- Gardezi, S. J. S., Faye, I., Bornot, J. M. S., Kamel, N., & Hussain, M. (2017). Mammogram classification using dynamic time warping. *Multimedia Tools and Applications*, 1-22. <https://doi.org/10.1007/s11042-016-4328-8>
- Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. MIT Press. <http://www.deeplearningbook.org>
- Görmez, Y., Sabzekekar, M., & Aydın, Z. (2021). IGPREP: Combination of convolutional neural and graph convolutional networks for protein secondary structure prediction. *Proteins: Structure, Function, and Bioinformatics*, 89(10), 1277-1288. <https://doi.org/10.1002/prot.26149>
- Griol, D., Baena, I., Molina, J. M., & de Miguel, A. S. (2014). A Multimodal Conversational Agent for Personalized Language Learning. İçinde C. Ramos, P. Novais, C. E. Nihan, & J. M. Corchado Rodríguez (Ed.), *Ambient Intelligence—Software and Applications* (ss. 13-21). Springer International Publishing. [https://doi.org/10.1007/978-3-319-07596-9\\_2](https://doi.org/10.1007/978-3-319-07596-9_2)
- Hill, J. R., & Jolly, N. A. (2012). Salary Distribution and Collective Bargaining Agreements: A Case Study of the NBA. *Industrial Relations: A Journal of Economy and Society*, 51(2), 342-363. <https://doi.org/10.1111/j.1468-232X.2012.00680.x>
- Hopfield, J. J. (1988). Artificial neural networks. *IEEE Circuits and Devices Magazine*, 4(5), 3-10. <https://doi.org/10.1109/101.8118>
- Hsieh, T.-C., Wang, T.-I., Su, C.-Y., & Lee, M.-C. (2012). A Fuzzy Logic-based Personalized Learning System for Supporting Adaptive English Learning. *Journal of Educational Technology & Society*, 15(1), 273-288.
- Hsu, C.-K., Hwang, G.-J., & Chang, C.-K. (2013). A personalized recommendation-based mobile learning approach to improving the reading performance of EFL students. *Computers & Education*, 63, 327-336. <https://doi.org/10.1016/j.compedu.2012.12.004>
- Janssens, A. C. J. W., Duijn, V., & M, C. (2008). Genome-based prediction of common diseases: Advances and prospects. *Human Molecular Genetics*, 17(R2), R166-R173. <https://doi.org/10.1093/hmg/ddn250>
- Jeong, H.-Y., Choi, C.-R., & Song, Y.-J. (2012). Personalized Learning Course Planner with E-learning DSS using user profile. *Expert Systems with Applications*, 39(3), 2567-2577. <https://doi.org/10.1016/j.eswa.2011.08.109>
- Keras: The Python deep learning API. (t.y.). Geliş tarihi 10 Temmuz 2021, gönderen <https://keras.io/>
- Kim, J., Kim, J., Thu, H. L. T., & Kim, H. (2016). Long Short Term Memory Recurrent Neural Network Classifier for Intrusion Detection. *2016 International Conference on Platform Technology and Service (PlatCon)*, 1-5. <https://doi.org/10.1109/PlatCon.2016.7456805>
- Koesdwiady, A., Soua, R., & Karray, F. (2016). Improving Traffic Flow Prediction With Weather Information in Connected Cars: A Deep Learning Approach. *IEEE Transactions on Vehicular Technology*, 65(12), 9508-9517. <https://doi.org/10.1109/TVT.2016.2585575>
- Kruppa, J., Ziegler, A., & König, I. R. (2012). Risk estimation and risk prediction using machine-learning methods. *Human Genetics*, 131(10), 1639-1654. <https://doi.org/10.1007/s00439-012-1194-y>
- Kumar, K., Parida, M., & Katiyar, V. K. (2013). Short Term Traffic Flow Prediction for a Non Urban Highway Using Artificial Neural Network. *Procedia - Social and Behavioral Sciences*, 104, 755-764. <https://doi.org/10.1016/j.sbspro.2013.11.170>
- li, W., & Liu, Z. (2011). A method of SVM with Normalization in Intrusion Detection. *Procedia Environmental Sciences*, 11, 256-262. <https://doi.org/10.1016/j.proenv.2011.12.040>
- Mesleki Yönelim Envanteri (Prof. Dr. Muharrem KEPÇEOĞLU). (2019, Aralık 4). *Eduolog.com*. <https://www.eduolog.com/mesleki-yonelim-envanteri-prof-dr-muharrem-kepceoglu/>
- Oktaç, A. (2015). Foreign Language Teaching: A Problem in Turkish Education. *Procedia - Social and Behavioral Sciences*, 174, 584-593. <https://doi.org/10.1016/j.sbspro.2015.01.587>
- Peng, H., Ma, S., & Spector, J. M. (2019). Personalized adaptive learning: An emerging pedagogical approach enabled by a smart learning environment. *Smart Learning Environments*, 6(1), 9. <https://doi.org/10.1186/s40561-019-0089-y>
- Petersen, S. A., Markiewicz, J.-K., & Bjørnebekk, S. S. (2009). Personalized and contextualized language learning: Choose when, where and what. *Research and Practice in Technology Enhanced Learning*, 04(01), 33-60. <https://doi.org/10.1142/S1793206809000635>
- Peterson, L. E. (2009). K-nearest neighbor. *Scholarpedia*, 4(2), 1883. <https://doi.org/10.4249/scholarpedia.1883>
- Prabusankarlal, K. M., Thirumoorthy, P., & Manavalan, R. (2015). Assessment of combined textural and morphological features for diagnosis of breast masses in ultrasound. *Human-Centric Computing and Information Sciences*, 5(1), 12. <https://doi.org/10.1186/s13673-015-0029-y>

- Precision and recall*. (2017). [https://en.wikipedia.org/wiki/Precision\\_and\\_recall](https://en.wikipedia.org/wiki/Precision_and_recall)
- Qoussini, A. E. M., & Bt Jusoh, Y. Y. (2014). A Review on Personalization and Agents Technology in Mobile Learning. *2014 International Conference on Intelligent Environments*, 260-264. <https://doi.org/10.1109/IE.2014.49>
- Scikit-learn: Machine learning in Python—Scikit-learn 0.24.2 documentation*. (t.y.). Geliş tarihi 10 Temmuz 2021, gönderen <https://scikit-learn.org/stable/>
- Shone, N., Ngoc, T. N., Phai, V. D., & Shi, Q. (2018). A Deep Learning Approach to Network Intrusion Detection. *IEEE Transactions on Emerging Topics in Computational Intelligence*, 2(1), 41-50. <https://doi.org/10.1109/TETCI.2017.2772792>
- Shuib, M., Abdullah, A., Azizan, S. N., & Gunasegaran, T. (2015). Designing an Intelligent Mobile Learning Tool for Grammar Learning (i-MoL). *International Journal of Interactive Mobile Technologies (IJIM)*, 9(1), 41. <https://doi.org/10.3991/ijim.v9i1.4238>
- Skopt module*. (2019). skopt module. <https://scikit-optimize.github.io/>
- Stergiou, C., & Psannis, K. E. (2017). Recent advances delivered by Mobile Cloud Computing and Internet of Things for Big Data applications: A survey. *International Journal of Network Management*, 27(3), e1930. <https://doi.org/10.1002/nem.1930>
- Thacker, R. A. (1995). Gender, influence tactics, and job characteristics preferences: New insights into salary determination. *Sex Roles*, 32(9), 617-638. <https://doi.org/10.1007/BF01544215>
- Traxler, J. (2007). Current State of Mobile Learning. *Mobile Learning: Transforming the Delivery of Education and Training*.
- Troussas, C., Chrysafiadi, K., & Virvou, M. (2018). Machine Learning and Fuzzy Logic Techniques for Personalized Tutoring of Foreign Languages. İçinde C. Penstein Rosé, R. Martínez-Maldonado, H. U. Hoppe, R. Luckin, M. Mavrikis, K. Porayska-Pomsta, B. McLaren, & B. du Boulay (Ed.), *Artificial Intelligence in Education* (ss. 358-362). Springer International Publishing. [https://doi.org/10.1007/978-3-319-93846-2\\_67](https://doi.org/10.1007/978-3-319-93846-2_67)
- Wei Yang, Kuanquan Wang, & Wangmeng Zuo. (2011). A fast and efficient nearest neighbor method for protein secondary structure prediction. *2011 3rd International Conference on Advanced Computer Control*, 224-227. <https://doi.org/10.1109/ICACC.2011.6016402>
- Wu, T.-T., Huang, Y.-M., Chao, H.-C., & Park, J. H. (2014). Personalized English reading sequencing based on learning portfolio analysis. *Information Sciences*, 257, 248-263. <https://doi.org/10.1016/j.ins.2011.07.021>
- Xu, B., Zhu, X., & Zhu, H. (2019). An Efficient Indoor Localization Method Based on the Long Short-Term Memory Recurrent Neuron Network. *IEEE Access*, 7, 123912-123921. <https://doi.org/10.1109/ACCESS.2019.2937831>
- Yao, C.-B. (2017). Constructing a User-Friendly and Smart Ubiquitous Personalized Learning Environment by Using a Context-Aware Mechanism. *IEEE Transactions on Learning Technologies*, 10(1), 104-114. <https://doi.org/10.1109/TLT.2015.2487977>
- Yarandi, M., Jahankhani, H., Dastbaz, M., & Tawil, A. R. (2011). *PERSONALISED MOBILE LEARNING SYSTEM BASED ON ITEM RESPONSE THEORY*. 8.
- Yeung, C. Y., & Lee, J. (2018). Personalized Text Retrieval for Learners of Chinese as a Foreign Language. *Proceedings of the 27th International Conference on Computational Linguistics*, 3448-3455. <https://aclanthology.org/C18-1292>
- Zhou, J., & Wang, J. (2016). Unsupervised fabric defect segmentation using local patch approximation. *The Journal of The Textile Institute*, 107(6), 800-809. <https://doi.org/10.1080/00405000.2015.1131440>