



Türkiye'de mutluluk düzeyini etkileyen faktörlerin makine öğrenme ve nitelik seçme algoritmaları ile belirlenmesi

Identifying the factors affecting the level of happiness in Turkey using machine learning and feature selection algorithms

İncilay Yıldız¹ , Fatih Abut^{2,*} 

¹ Çukurova Üniversitesi, Mühendislik Fakültesi, Endüstri Mühendisliği Bölümü, 01330, Adana, Türkiye

² Çukurova Üniversitesi, Mühendislik Fakültesi, Bilgisayar Mühendisliği Bölümü, 01330, Adana, Türkiye

Özet

Mutluluk düzeyi, mutluluk algısının düzeyini açıklayan ve yaşam memnuniyeti ölçmeye yarayan önemli bir parametredir. Bu çalışmada, Türkiye'deki mutluluk düzeyini etkileyen faktörleri belirlemek üzere Relief-F nitelik seçme algoritması ile birleştirilmiş Destek Vektör Makinesi (SVM), Çok Katmanlı Sinir Ağı (MLP) ve Ağaç Arttırma (TB) tabanlı yeni modeller önerilmiştir. Kullanılan veri seti Türkiye İstatistik Kurumu (TÜİK) tarafından 2015 yılında il düzeyinde yapılan İllerde Yaşam Endeksi (İYE) araştırması sonucu oluşturulmuştur. Tahmin modellerinin performansı, 10-katlı çapraz doğrulama kullanılarak, Kök Ortalama Kare Hatası (Root Mean Square Error - RMSE) ve R-kare (R^2) değerlerinin hesaplanmasıyla değerlendirilmiştir. Elde edilen sonuçlara göre sağlık hizmetlerinden memnuniyet oranı, sağlığından memnuniyet oranı ile orta ve üstü gelir grubundaki hanelerin oranı Türkiye'deki mutluluk düzeyi tahmininde önemli rol oynayan değişkenler olarak belirlenmiştir. Ayrıca SVM yönteminin mutluluk düzeyi tahmininde etkin bir yöntem olduğu ve tüm SVM tabanlı tahmin modellerinin MLP ve TB tabanlı modellere göre daha düşük RMSE ve daha yüksek R^2 değerleri ürettiği gözlemlenmiştir.

Anahtar kelimeler: Makine öğrenmesi, Nitelik seçme, Mutluluk düzeyi, Tahmin

1 Giriş

Mutluluk, bir bireyin kendi yaşamının genel kalitesini bir bütün olarak olumlu bir şekilde değerlendirme derecesidir. Genellikle yaşamda nihai bir hedef olarak kabul edilir ve neredeyse herkes mutlu olmak ister. Mutluluk; gelir, işgücü piyasası durumu, iş özellikleri, sağlık, boş zaman, aile, sosyal ilişkiler, güvenlik, özgürlük, ahlaki değerler ve diğerleri dahil olmak üzere birçok etkene bağlıdır [1]. TÜİK tarafından mutluluk "acı, keder ve ıstırahın yokluğu ve bunların yerine sevinç, neşe ve tatmin duygularının varlığıyla karakterize edilen durum; hayattan genel olarak memnun olma hali" şeklinde ifade edilmiştir [2].

Mutluluk verilerinin en temel kullanımı, bir ülke veya belirli bir sosyal grup içindeki görünen yaşam kalitesini tahmin etmektir. Bu genellikle bir sosyal problemin veya

Abstract

Happiness level is an important parameter that explains the level of perception of happiness and helps to measure life satisfaction. In this study, new models based on Support Vector Machine (SVM), Multilayer Neural Network (MLP) and Tree Boost (TB) individually combined with Relief-F feature selector algorithm were proposed to determine the factors affecting the level of happiness in Turkey. The utilized data set was created by the Turkish Statistical Institute (TÜİK) in 2015 as a result of the Provincial Life Index survey. The performance of the prediction models was evaluated by calculating the Root Mean Square Error (RMSE) and R-squared (R^2) values using 10-fold cross validation. According to the results obtained, satisfaction with health services, health satisfaction rate and the ratio of households in the middle- and higher-income group were identified as variables that play an important role in the prediction of happiness level in Turkey. In addition, it has been observed that SVM is an effective method in predicting happiness level and all SVM-based prediction models consistently produce lower RMSE and higher R^2 values than MLP-based and TB-based models.

Keywords: Machine learning, Feature selection, Happiness level, Prediction

konunun kapsamını değerlendirmek ve olası politika müdahalelerini önermek için yapılır. Yüksek memnuniyet, ilgili popülasyondaki yaşam kalitesinin iyi olduğunu gösterirken düşük memnuniyet, ciddi eksikliklere işaret etmektedir [3].

Birleşmiş Milletler 2019 Yılı Dünya Mutluluk Raporu'na göre; dünyanın en mutlu ülkesi Finlandiya en mutsuz ülkesi Güney Sudandır. Toplam 156 ülkenin yer aldığı araştırmada Türkiye, 2017 yılında 69. sıradayken geçtiğimiz yıl 74'üncü sırada ve bu yıl 79. sırada yer almıştır [4]. YMA'nın 2019 yılı sonuçlarına göre Türkiye'de genel mutluluk düzeyi ve mutluluk kaynağı olan değerlerin yüzdeleri Tablo 1'de verilmiştir [5].

* Sorumlu yazar / Corresponding author, e-posta / e-mail: fabut@cu.edu.tr (F. Abut)

Geliş / Received: 02.02.2021 Kabul / Accepted: 26.07.2021 Yayımlanma / Published: 14.01.2022

doi: 10.28948/ngumuh.873199

Tablo 1. YMA 2019 sonuçları

YMA 2019	Değer Tanımı	%
Genel Mutluluk Düzeyi	Çok Mutlu	6.61
	Mutlu	45.74
	Orta	34.58
	Mutsuz	9.93
Mutluluk Kaynağı Olan Değerler	Çok Mutsuz	3.13
	Başarı	9.17
	İş	2.13
	Sağlık	69.86
	Sevgi	14.32
Para	4.16	
Diğer	0.35	

Mutluluk düzeyi; toplumun refah içinde yaşaması, motivasyonu ile doğrudan ilişkili olduğu için kişinin verimliliği ve başarısı için de tetikleyici bir unsurdur. Başarılı ve mutlu toplumlar iyi bir yaşam kalitesine sahiptir. Yaşam kalitesi mutluluk düzeyi ile ölçüldüğü için mutluluk düzeyini tahmin etmek önemlidir.

Bu çalışma kapsamında, Türkiye'de mutluluk düzeyini etkileyen faktörleri belirlemek üzere Relief-F nitelik seçici algoritması ile birleştirilmiş Destek Vektör Makinesi (SVM), Çok Katmanlı Sinir Ağı (MLP) ve TB tabanlı çeşitli tahmin modelleri geliştirilmiştir. Elde edilen sonuçlar, 10-katlı çapraz doğrulama kullanılarak, Kök Ortalama Kare Hatası (Root Mean Square Error - RMSE) ve R-kare (R^2) değerlerinin hesaplanmasıyla değerlendirilmiş ve karşılaştırılmıştır.

Bölüm 2'de literatürde bulunan ilgili çalışmalar özetlenmiştir. Bölüm 3'te veri setini oluşturma sürecinden bahsedilmiş ve metodoloji hakkında bilgiler verilmiştir. Bölüm 4'te çalışma ile ilgili sonuçlar ve tartışma sunulmuştur. Bölüm 5, makaleyi özet ve öneriler ile sonlandırmaktadır.

2 İlgili çalışmalar

Literatürde mutluluk düzeyi ve yaşam memnuniyeti ile ilgili birçok çalışma mevcuttur. Sibel Selim'in [6] 2008 yılında yapmış olduğu araştırma, Türkiye'deki yaşam doyumunu ve mutluluğunu incelemeyi amaçlamaktadır. Türkiye için öznel iyi oluş ile ilgili önceki araştırmalar hem mutluluk hem de yaşam doyumunu göz önünde bulundurarak genişletilmiştir. Bu çalışmada, Türk halkı arasında bireysel mutluluğu ve yaşam memnuniyetini şekillendiren faktörler, sosyokültürel ve politik değişimin temsili bir anketi olan Dünya Değerler Araştırması (WVS) kullanılarak belirlenmiştir. Tahmin modellerinde sıralı logit modeli kullanılırken modelleri yorumlamak için ayrık değişim kullanılmıştır. Ayrıca tahmin modelleri oluşturulurken yaş, din, öznel sağlık, cinsiyet, eğitim düzeyi, medeni durum, çocuk sayısı, istihdam durumu ve gelir değişkenleri kullanılmıştır. Sağlık, yanıtlayanların beş puanlık bir ölçekte kendi kendine derecelendirmesiyle ölçülürken, gelir on puanlık bir ölçek olarak ölçülür. Bu değişkenlere ek olarak modellerde ailenin önemi, arkadaşlar, boş zaman, siyaset, hayatta çalışma, milliyetten gurur, insana güven ve siyasi yönelim değişkenleri kullanılmıştır. Yaş değişkeni 15–24, 25–34, 35–44 ve 45+ olmak üzere dört yaş grubuna ayrılırken, eğitim değişkeni alt, orta ve üst eğitim

olmak üzere üç düzeyde kategorize edilmiştir. Ek olarak, medeni durum değişkenine ilişkin kategoriler evli, hiç evlenmemiş ve boşanmış, ayrılmış veya duldur. Ayrıca, tam zamanlı, yarı zamanlı, serbest meslek sahibi, emekli, ev hanımı, öğrenciler, işsizler ve diğerleri istihdam türü değişkenleridir. Son olarak, gelir seviyeleri değişkeni sekiz gruba ayrılmıştır. Tahmin modellerinin ölçümünde tahmin katsayısı ve marjinal etki kullanılmıştır. Elde edilen bazı sonuçlar, olumsuz bir yaş etkisi, gelir ve sağlık durumunun olumlu etkileri ve işsizliğin olumsuz etkisi gibi tipik bulgularla benzerdir. Beklentilerin aksine, orta eğitim kadınlarda yaşam doyumunu olumsuz yönde doğrudan etkiliyor ve yaşam doyumunu modelinde üst eğitim düzeyi önemsiz. Farklı yıllar boyunca mutluluk ve yaşam doyumunu ilişkilerini karşılaştıran bu araştırma, Türk halkının algılanan yaşamlarını iyileştirmek için politika yapıcılara ve profesyonellere içgörü sağlamayı amaçlamaktadır.

Gunvor Marie Dyrdal ve arkadaşlarının [7] 2010 yılında yapmış oldukları çalışmada, genel yaşam memnuniyeti ve ilişki tatmini arasındaki bağlantı, Norveç Halk Sağlığı Enstitüsü tarafından yürütülen Norveç Anne ve Çocuk Kohort Çalışmasından (MoBa) elde edilen veriler kullanılarak anneler arasında boylamsal olarak araştırılmıştır. Mevcut çalışmada, Norveç Anne ve Çocuk Kohort Çalışması 1999 yılında başlayan ve 2008 yılına kadar 100,000 hamile kadının işe alındığı bir hamilelik kohortudur. Norveç'teki hamile kadınların çoğu çalışmaya katılmaya davet edilmiştir ve yaklaşık %45 yanıt oranıyla katılım sağlanmıştır. Veriler hamilelik sırasında iki kez ve doğum sonrası 6. ve 36. aylarda toplanmıştır. Verilerin ortalama ve standart sapması hesaplanmıştır, analiz yöntemi olarak ise yapısal eşitlik modellemeleri kullanılmıştır. Bulgular, çapraz gecikmeli uzunlamasına modeli kullanan yapısal eşitlik modellemesi hem genel yaşam memnuniyeti hem de ilişki tatmini için çapraz kavramlar arası çapraz zaman etkilerini kanıtlamıştır. Elde edilen sonuçlara göre, doğumdan hemen sonra ilişki tatmininin azaldığı ve genel yaşam memnuniyetinin başlangıçta bir artış göstermesine rağmen doğum sonrası bir düşüş gösterdiği tespit edilmiştir. Sonuçlar, genel yaşam memnuniyeti ve ilişki tatmini seviyelerinin, kesitsel ilişkilerinin olduğu gibi, zaman içinde oldukça istikrarlı olduğunu göstermiştir. Tatmin edici bir romantik ilişkiye sahip olmak, gelecekteki yaşam doyumunu korumak ve artırmak için önemlidir.

İraj Mirkhan ve arkadaşlarının [8] 2014 yılında yapmış oldukları çalışmada, Mutluluk düzeyi, öğretmenlerin eğitim kalitesinin artırılmasına ve öğrenciler için uygun bir ortam yaratılmasına yardımcı olmaktadır. Bu çalışmada, Urmiye şehrinin kadın öğretmenlerinin duygusal zekâ, mutluluk ve dini tutumlara dayalı olarak mutluluk düzeyinin tahminini araştırmak amaçlanmıştır. Çalışma kapsamında, Urmiye liselerinden rastgele seçilen 262 öğretmen katılımcıya Khodayarifard dini tutum anketini, Oxford Mutluluk, Duygusal zekâ Bar-On Envanteri ve Diener Yaşam Ölçeği ile memnuniyet anketi yapılmıştır. Araştırma değişkenleri olarak yaş, öğretim geçmişi, yaşam doyumunu, dini tutumlar, mutluluk ve duygusal zekâ kullanılmıştır. Anketten elde edilen verilerin ortalama ve standart sapma değerleri, Pearson korelasyon ve regresyon testi kullanılarak SPSS yazılımı ile

analiz edilmiştir. Sonuçlar incelendiğinde, kadın öğretmenlerde duygusal zekâ ile yaşam doyumu ve mutluluk ile yaşam doyumu arasında pozitif bir ilişki olduğu ancak dini tutum ile yaşam doyumu arasında bir ilişki olmadığı gözlemlenmiştir. Sonuçlar ayrıca mutluluğun yaşam doyumunu tahmin etmekte tek belirleyici olduğunu göstermektedir. Çalışmadan elde edilen bulgular, duygusal zekâ ve mutluluk bileşenlerinin kadın öğretmenlerin yaşam doyumlarında önemli bir rol oynadığını göstermektedir.

Noei Iran ve Mahmoud Alilou'nun [9] 2015 yılında yapmış oldukları çalışmada bağlanma stilleri, farklı önemli psikolojik yapıların tahmin edilmesinde önemli rol oynayan en önemli psikolojik değişkenler arasındadır. Bu çalışmanın amacı, Tebriz-İran'daki evli çiftlerde bağlanma stillerine göre yaşam kalitesi ve mutluluğu tahmin etmektir. 1393 yılında Tebriz şehrinde yaşayan 100 evli çift (50 kadın ve 50 erkek) katılımcıya, Dünya Sağlık Örgütü Yaşam Kalitesi (WHOQOL BREF), Oxford Mutluluk Anketi (OHQ) ve Hazan ve Shaver Yetişkin Bağlanma ölçeğinin kısaltılmış formu ile anket uygulanmıştır. Verilerin geçerliliği ve güvenilirliği Pearson korelasyon katsayısı ve çoklu regresyon kullanılarak SPSS programı ile analiz edilmiştir. Elde edilen regresyon analizi sonuçları bağlanma stilinin yaşam kalitesi ve mutluluğu tahmin edebilir olduğunu ortaya koymuştur. Çalışmanın sonucu; Bağlanma stilleri insanların yaşam kalitesini ve mutluluğunu tahmin etmede önemli bir rol oynamaktadır, bu nedenle yaşam kalitesini ve mutluluğunu artırmaya yönelik programların geliştirilmesinde katılımcıların bağlanma stilleri dikkate alınmalıdır.

Natasha Jaques ve arkadaşlarının [10] 2015 yılında yapmış oldukları çalışmada, öğrencilerin mutluluğunu modellemek için, her biri bir ay boyunca izlenen lisans öğrencilerinden toplanan verilere makine öğrenimi yöntemleri uygulanmıştır. Bunlar Destek Vektör Makineleri, Rasgele Ormanlar, Sinir Ağları, Doğrusal Regresyon, k-En Yakın Komşu arama (kNN), Naive Bayes ve Adaboost yöntemleridir. En iyi sınıflandırıcıyı bulduktan sonra, sınıflandırıcının parametre alanı aranmış ve doğrulama setindeki performansını optimize eden parametreler seçilmiştir. Toplanan veriler arasında fizyolojik sinyaller, konum, akıllı telefon günlükleri ve davranışsal sorulara anket yanıtları bulunmaktadır. Katılımcılar her gün stres, sağlık ve mutluluk gibi önlemler hakkındaki durumlarını bildirmişlerdir. Mutluluk ve depresyon arasındaki ilişki, mutluluğu modellemek, depresyon riski taşıyan bireyleri tespit edilmesine ve katılımcılara yardım edecek müdahaleleri önermeye yardımcı olabildiği için önemlidir. Ayrıca bu çalışma davranışsal faktörlerin (uyku ve sosyal aktivite gibi) mutluluğu nasıl olumlu ve olumsuz etkilediğiyle de ilgilenmektedir. Gauss Karışım Modelleri ve topluluk sınıflandırması dahil olmak üzere çeşitli makine öğrenimi ve özellik seçme teknikleri karşılaştırılmıştır. Sonuç olarak, test verilerinde bildirilen mutluluğun %70 sınıflandırma doğruluğuna ulaşılmıştır.

Asena Caner'in [11] 2016 yılında yapmış olduğu çalışmada, Türkiye'de son yıllardaki mutluluk ve yaşam doyumunun belirleyicilerini incelemeyi amaçlanmıştır. 2007 ve 2011 yıllarında ulusal düzeyde temsili hane halkı üzerinde

yapılan Türkiye Yaşam Memnuniyeti Anketi ve Türkiye için Dünya Değerler Anketi'nden elde edilen mikro verilere dayalı regresyon analizlerini kullanılmıştır. Bu çalışmada, iki anket tarafından toplanan veriler ve sonuç değişkeninin korelasyonları kullanılmıştır. Sonuçlardan bazıları hem literatürdeki öncekilerle hem de bu çalışmada kullanılan veri kümeleriyle tutarlı olsa dahi, bazı sonuçlar zaman periyoduna, kontrol değişkenleri kümesine ve kullanılan örneğe bağlıdır; bu nedenle bulguların verimliliği şüphe uyandırmaktadır. Bu çalışmada, doğru bir karşılaştırma yapmak için bireyin mutluluk düzeyini belirlerken, bireyin demografik ve ekonomik özelliklerinin bir fonksiyonu olarak geliştirilen mutluluk denklemleri kullanılarak tahmin işlemi yapılmıştır. Tahminler marjinal etki ve standart hata metrikleri kullanılarak ölçülmüştür. Tahminler sırasında benzer bulunan değişkenler işsizlik durumu, medeni durum, gelir ve cinsiyettir. Yaş, mutlak gelir ve eğitim tahminlerinde farklılıklar görülmektedir. Çalışmanın sonucunda, mutluluğu hangi faktörlerin belirlediği sorusuna dair daha detaylı verilerle çalışılması gerektiği vurgulanmıştır. Ayrıca, veri kalitesini iyileştirmenin yolları için öneriler yapılmıştır.

Malathie Dissanayake ve arkadaşlarının [12] 2017 yılında yapmış oldukları çalışmanın amacı, bireylerin yaşam algılarının öznel mutluluk ve yaşam doyumu üzerindeki etkisini incelemektir. Katılımcılar 18-28 yaşları arasında (yaş ortalaması 23 yaşın standart sapması 1.60) %62'si erkek, %38'i kadından oluşan 200 Hintli bireyden oluşmaktadır. Katılımcılara uygulanan anket; yaş, cinsiyet, medeni durum, din, etnik köken, eğitim, istihdam, yaşam alanı ile ilgili maddeleri içeren sosyodemografik soruları içermektedir. Güvenirlik testi çalışmasında kullanılan değişkenler, yaşamdaki amaç, olumlu duygulanım, öznel mutluluk ve yaşam doyumudur. Regresyon analizi ve güvenirlik testi ile elde edilen sonuçlar, hayattaki amacın kişinin mutluluğunu ve yaşam doyumunu önemli ölçüde tahmin ettiğini ortaya koymuştur. Ayrıca olumlu duygulanım, yaşamdaki amaç ile öznel mutluluk arasındaki ilişkiye aracılık etmektedir. Yaşam amacı ile yaşam doyumunu arasındaki bağlantının olumlu duygulanım aracılığıyla gerçekleştiği ortaya koymuştur. Bu durum, hayatlarını daha anlamlı ve maksatlı algılayan bireylerin daha olumlu bir etki, dolayısıyla mutlu ve tatmin hissettiklerini göstermektedir.

Bu çalışmada, literatürden farklı olarak daha güncel yaşam memnuniyeti verileri kullanılmıştır. Ayrıca Türkiye'nin il bazında yapılan YMA çalışmasının verilerini kullanarak mutluluk düzeyini yoğun istatistiksel hesaplamalara gerek kalmadan, makine öğrenme ve nitelik seçme algoritmaları kullanarak etkin bir şekilde tahmin edebilen modeller önerilmiştir.

3 Materyal ve metod

Bu çalışmada kullanılan veri seti, TÜİK tarafından yapılan 2015 yılı İYE çalışmasının verileridir. Bu veriler, endeks objektif ve subjektif göstergelerinden oluşmaktadır. Subjektif göstergelerinin temel veri kaynağı il düzeyinde uygulanması planlanan Yaşam Memnuniyeti Araştırmasının (YMA) sonuçlarıdır. Objektif göstergeler ise il düzeyinde TÜİK verileri ve diğer kurum ya da kuruluşlardan derlenen verilerden oluşmaktadır [13]. YMA, 2015 yılı 4500 cevaplı

örnek hanede 9397 cevaplı bireyle gerçekleştirilen araştırmasından elde edilmiştir. YMA'nın amacı TÜİK tarafından "Türkiye'deki bireylerin öznel mutluluk algısı, sağlık, sosyal güvenlik, örgün eğitim, çalışma hayatı, gelir, kişisel güvenlik ve adalet hizmetleri, kişisel gelişim gibi temel yaşam alanlarındaki memnuniyetleri ölçmek ve bunların zaman içindeki değişimi takip etmek" şeklinde açıklanmıştır [14]. YMA'nın kapsamı, Türkiye Cumhuriyeti sınırları içinde bulunan hanelerde yaşayan 18 ve daha yukarı yaşta T.C. vatandaşları ile yabancı uyruklu kişilerdir. Araştırmada, kurumsal nüfus (üniversite öğrenci yurdu, huzurevi, bakımevi, ceza ve tutukevi, ıslahevi, yetiştirme yurdu, askeri birlik ve kışla vb.) kapsam dışı tutulmuştur. YMA ilk kez 2003 yılı Kasım ayında hane halkı Bütçe Anketi'ne ek bir modül olarak gerçekleştirilmiştir. Araştırma 2003 yılından itibaren her yıl düzenli olarak yapılmaktadır [15].

Mutluluk düzeyini tahmin etmek üzere dört farklı veri grubuna ait çeşitli değişkenler kullanılmıştır. Birinci veri grubu çalışma hayatına ilişkin olarak istihdam oranı (ISHO), işsizlik oranı (ISSO), ortalama günlük kazanç (OGK) ve işinden memnuniyet oranı (IMO) değişkenlerini içermektedir. İkinci veri grubu gelir ve servet düzeyi ile ilgili olarak kişi başına düşen tasarruf mevduatı (KTM), orta ve üstü gelir grubundaki hanelerin oranı (OUG) ve temel ihtiyaçlarını karşılayamadığını beyan eden hanelerin oranı (TIK) değişkenlerini içermektedir. Üçüncü veri grubu sağlık konuları ile alakalı olarak bebek ölüm hızı (BOH), doğuştan beklenen yaşam süresi (DBY), hekim başına düşen müracaat sayısı (HDM), sağlığından memnuniyet oranı (SMO) ve kamunun sağlık hizmetlerinden memnuniyet oranı (KSH) değişkenlerinden oluşmaktadır. Dördüncü veri grubu ise eğitim konularına ilişkin okul öncesi eğitimde (3-5 yaş) net okullaşma oranı (OOE), TEOG sistemi yerleştirmeye esas puan ortalaması (TEOG), YGS puan ortalaması (YGS), fakülte veya yüksekokul mezunlarının oranı (FYM), kamunun eğitim hizmetlerinden memnuniyet oranı (KEM) değişkenlerini kapsamaktadır. Ayrıca ID değişkeni Türkiye'nin 81 ilini içermekte olup tüm veri gruplarında ortaktır. Veri setinin betimsel istatistiği Tablo 2'de gösterilmiştir.

Relief-F algoritmasının bir sıralayıcı arama yöntemi ile çalıştırılmasıyla veri seti içerisindeki her bir tahmin değişkeninin önem puanı hesaplanmış ve değişkenler puanlarına uygun olarak en önemliden en önemsiz doğru sıralanmıştır. Relief-F tarafından hesaplanan tahmin değişkenlerinin puanları ve önem sıraları Tablo 3'te gösterilmiştir. Değişkenler arasındaki korelasyon ilişkilerini incelemek üzere her aşamada en düşük puana sahip değişken iteratif olarak tüm değişkenler kümesinden çıkarılarak toplamda 18 farklı mutluluk düzeyi tahmin modeli oluşturulmuştur. Daha detaylı olarak, ilk model çalışmada kullanılan tüm girdi değişkenlerin tam setini içermektedir. İkinci model, en düşük Relief-F sıralamasına sahip olan HDM dışında birinci modelin içerdiği aynı değişken kümesi kullanılarak oluşturulmuştur. Benzer şekilde, üçüncü model ikinci en düşük Relief-F sıralamasına sahip olan ID dışında ikinci modelin içerdiği aynı değişken grubunu içermektedir. Bu metodoloji, en yüksek puana sahip tek bir değişken

(KSH) son modeli oluşturana kadar tekrarlanmıştır. Tahmin modelleri ve her modelin içerdiği tahmin değişkenleri Tablo 4'te gösterilmiştir.

Tahmin modellerini geliştirmek için SVM, MLP ve TB olmak üzere 3 farklı makine öğrenme yöntemi kullanılmıştır.

Tablo 2. Veri setinin betimsel istatistiği

Değişkenler	Min.	Mak.	Ortalama	Standart Sapma
ID	1.00	81.00	41.00	23.52
ISHO	27.80	59.10	46.21	6.20
ISSO	4.20	23.40	8.80	3.87
OGK	46.90	85.60	57.68	6.57
IMO	64.00	91.60	78.78	6.47
KTM	616.00	18.131.00	4.342.11	2.821.42
OUG	16.30	58.90	34.38	8.63
TIK	32.80	75.00	50.95	10.18
BOH	5.30	25.70	10.99	3.39
DBY	75.00	80.50	78.13	1.03
HDM	2.763.00	8.067.00	5.834.38	1.245.11
SMO	59.20	80.80	72.00	4.44
KSH	54.60	89.10	77.47	7.30
OOE	23.50	53.20	35.27	6.25
TEOG	215.30	338.00	295.93	27.67
YGS	178.60	207.90	197.64	5.51
FYM	8.60	22.70	13.12	2.34
KEM	48.20	88.90	74.09	8.45
MD	41.98	77.66	61.15	7.53

Tablo 3. Relief-F ile sıralanmış nitelikler

Nitelik	Relief-F Skoru	Önem Sırası
KSH	1.6 ± 0.66	1
SMO	1.7 ± 0.64	2
KEM	3.3 ± 1.27	3
ISSO	5.7 ± 2.19	4
TIK	5.8 ± 1.08	5
IMO	5.9 ± 2.07	6
ISHO	6.2 ± 1.78	7
OUG	7.5 ± 1.28	8
YGS	8.4 ± 2.33	9
TEOG	9.4 ± 1.28	10
OOE	11.7 ± 1.85	11
FYM	12.0 ± 1.10	12
DBY	13.5 ± 1.43	13
OGK	14.1 ± 1.45	14
KTM	14.5 ± 1.69	15
BOH	15.2 ± 0.98	16
ID	16.5 ± 1.20	17
HDM	18.0 ± 0	18

Tablo 4. Tahmin modellerinin içerdikleri değişkenler

Model No	İçerdiği Değişkenler
Model 1	KSH, SMO, KEM, ISSO, TIK, IMO, ISHO, OUG, YGS, TEOG, OOE, FYM, DBY, OGK, KTM, BOH, ID, HDM
Model 2	KSH, SMO, KEM, ISSO, TIK, IMO, ISHO, OUG, YGS, TEOG, OOE, FYM, DBY, OGK, KTM, BOH, ID
Model 3	KSH, SMO, KEM, ISSO, TIK, IMO, ISHO, OUG, YGS, TEOG, OOE, FYM, DBY, OGK, KTM, BOH
Model 4	KSH, SMO, KEM, ISSO, TIK, IMO, ISHO, OUG, YGS, TEOG, OOE, FYM, DBY, OGK, KTM
Model 5	KSH, SMO, KEM, ISSO, TIK, IMO, ISHO, OUG, YGS, TEOG, OOE, FYM, DBY, OGK
Model 6	KSH, SMO, KEM, ISSO, TIK, IMO, ISHO, OUG, YGS, TEOG, OOE, FYM, DBY
Model 7	KSH, SMO, KEM, ISSO, TIK, IMO, ISHO, OUG, YGS, TEOG, OOE, FYM
Model 8	KSH, SMO, KEM, ISSO, TIK, IMO, ISHO, OUG, YGS, TEOG, OOE
Model 9	KSH, SMO, KEM, ISSO, TIK, IMO, ISHO, OUG, YGS, TEOG
Model 10	KSH, SMO, KEM, ISSO, TIK, IMO, ISHO, OUG, YGS
Model 11	KSH, SMO, KEM, ISSO, TIK, IMO, ISHO, OUG,
Model 12	KSH, SMO, KEM, ISSO, TIK, IMO, ISHO
Model 13	KSH, SMO, KEM, ISSO, TIK, IMO
Model 14	KSH, SMO, KEM, ISSO, TIK
Model 15	KSH, SMO, KEM, ISSO
Model 16	KSH, SMO, KEM
Model 17	KSH, SMO
Model 18	KSH

SVM görüntü tanıma, konuşma tanıma, metin sınıflandırması, yüz algılama ve hatalı kart algılama gibi birçok problem çeşidine uygulanabilmektedir [16]. Bir SVM modelinin performansı, C değeri, çekirdek fonksiyonunun tipi ve ilgili parametreleri ve ϵ değeri gibi çeşitli parametrelere bağlıdır. Çekirdek fonksiyonu olarak radyal tabanlı fonksiyon (Radial Basis Function – RBF) seçilmiştir. RBF çekirdeği, γ fonksiyon parametresinin optimizasyonunu gerektirmektedir.

Bir SVM modeli oluşturmadan önce, belirli bir problem için hangi C, ϵ ve çekirdek parametre değerlerinin en uygun olduğu önceden bilinemez. Bu anlamda, test verilerinin minimum hata ile tahmin edilebilmesi için C, ϵ , ve çekirdek parametrelerinin optimal değerlerini bulmak amacıyla bir tür parametre arama ve optimizasyon işlemi yapılmalıdır. Izgara arama yöntemi, orta büyüklükteki problemler için C, ϵ ve çekirdek parametrelerinin optimal değerlerini bulmak için kullanılan etkili bir yöntemdir. Yöntemin çalışma prensibi, parametrelerin değerlerini bir dizi değer aracılığıyla önceden belirlenebilen adım boyutlarıyla değiştirmektir.

Geliştirilen tahmin modellerinin geliştirilebilmesini sağlamak için veri setini eğitim ve test setlerine bölmek üzere 10 kat çapraz doğrulama kullanılmıştır. SVM tabanlı bir modelin akış şeması Şekil 1'de 10 katlı çapraz doğrulamanın tek katı için verilmiştir. Başlangıçta, eğitim ve test alt kümelerindeki girdi değişkenler, daha büyük sayısal aralıklardaki girdi değişkenlerin daha küçük sayısal aralıklardakilere baskın çıkmasını önlemek için sıfır ortalama ve birim varyansa sahip olacak şekilde standardizasyon kullanılarak ön işleme tabi tutulmuştur. Daha sonra C, ϵ ve γ parametrelerinin optimal değerlerini bulmak için eğitim seti üzerinde 10 katlı çapraz doğrulama

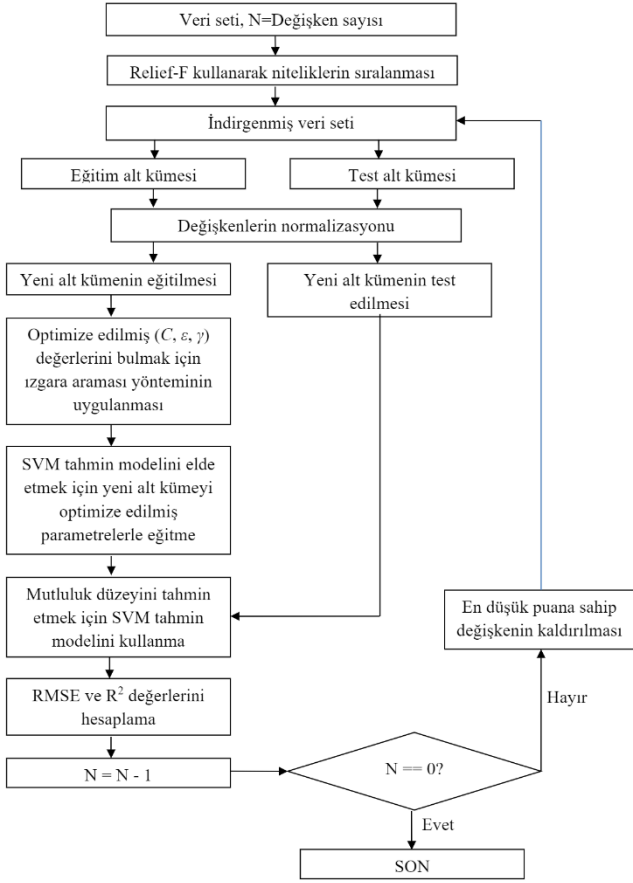
yapılmıştır. 10 katlı çapraz doğrulamada, eğitim seti rastgele on eşit büyüklükte alt kümeye bölünür. On alt kümeden, modeli test etmek için doğrulama verileri olarak tek bir alt küme tutulur ve kalan dokuz alt küme eğitim verileri olarak kullanılır. Çapraz doğrulama işlemi, doğrulama verileri olarak on alt kümenin her biri ile on kez tekrarlanmıştır. Daha sonra, tek bir tahmin üretmek için elde edilen on sonucun ortalaması alınmıştır. Eğitim setini eğitmek ve tahmin modelini oluşturmak için en düşük RMSE değerine yol açan üçlü parametre değerleri seçilmiştir.

Çok Katmanlı Algılayıcı İleri Beslemeli bir Yapay Sinir Ağı (YSA) yapısıdır. İleri beslemeli ağlarda veriler giriş katmandan çıkış katmana doğru iletilmektedir. Her bir katman yalnızca kendinden sonraki katman ile bağlanmaktadır. YSA sistemine gelen bilgiler öncelikle giriş katmanına sonrasında sırasıyla ara katmanlardan ve çıkış katmanından işlenerek geçer ve çıktı değerleri elde edilir [17].

Son olarak, ağaç yapılı regresyon yöntemleri kategorisinin popüler bir temsilcisi olan TB, performansını diğer iki kategoriyle karşılaştırmak için kullanılmıştır.

Tablo 5'te SVM, MLP ve TB tabanlı tahmin modellerinin sonuçlarını iyileştirmek için kullanılan parametre değerlerinin aralıkları verilmiştir.

Tahmin modellerinin performansı, Kök Ortalama Kare Hatası (Root Mean Square Error - RMSE) ve R^2 değerleri hesaplanarak değerlendirilmiştir. RMSE, kareler toplamının bölünmüş tahmin sayısının karekökü olarak tanımlanır ve bir model tarafından öngörülen değerler ile gerçekte gözlenen değerler arasındaki farkları ölçer. Düşük RMSE değeri, modelin daha doğru tahmin ürettiğini ifade eder. R^2 değeri ise bağımsız değişkenlerden tahmin edilebilen



Şekil 1. Tek bir kat için mutluluk düzeyini tahmin etmek için SVM tabanlı modelin akış şeması

bağımlı değişkendeki varyansın oranıdır. Sıfıra yakın bir R^2 değeri, gözlem verileriyle regresyon modeli arasında bir ilişki olmadığını gösterirken, bire yakın bir değer, mükemmel bir uyum gösterir. RMSE ve R^2 formülleri (1) ve (2)'de gösterilmiştir.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (Y - Y')^2} \quad (1)$$

$$R^2 = 1 - \frac{\sum (Y - Y')^2}{\sum (Y - \bar{Y})^2} \quad (2)$$

Denklem (1) ve (2)'de, Y gerçek mutluluk düzeyi değerini, Y' tahmin edilen mutluluk düzeyi değerini, \bar{Y} gerçek mutluluk düzeyi değerlerinin ortalamasını, n ise test edilen alt küme örneklerinin sayısını göstermektedir.

Tüm deneyler, çeşitli veri madenciliği tekniklerini barındıran ve ticari bir modelleme yazılımı olan DTREG¹ kullanılarak gerçekleştirilmiştir. DTREG, sınıflandırma ve regresyon modellerinin geliştirilmesine olanak sağlamaktadır ve gelecekteki gözlemler için değerleri tahmin edebilmektedir.

Tablo 5. Makine öğrenme yöntemlerinin parametre değerleri

Yöntem	Parametre	Değer
SVM	Cost (C)	[0.1-5000]
	Epsilon (ε)	[0.0001-0.6]
	Gamma (γ)	[0.001-50]
MLP	Gizli katmandaki nöron sayısı & adım sayısı	[2-80 & 1-5]
	Gizli katman aktivasyon fonksiyonu	Lojistik
	Çıktı katmanı aktivasyon fonksiyonu	Doğrusal
	Seride kullanılan maksimum ağaç sayısı	[50-600]
TB	Bireysel ağaçların derinliği	5
	Bölünme için minimum boyut düğümleri	10

4 Sonuçlar

Geliştirilen tüm tahmin modelleri üzerinden elde edilen RMSE ve R^2 değerleri Tablo 6'da gösterilmiştir. Elde edilen sonuçlara göre aşağıdaki tartışmalar yapılabilir:

- Relief-F algoritmasıyla en önemli değişken KSH, en etkisiz değişken ise HDM olarak sıralanmıştır.
- KSH, SMO, KEM, ISSO, TIK, IMO, ISHO, OUG, YGS ve TEOG değişkenlerini içeren Model 9 kullanılan yöntemden bağımsız olarak en düşük RMSE ve en yüksek R^2 değerlerini üretmiştir. Eğitim aşamasında Model 9'a ait gerçek ve tahmin edilen mutluluk düzeyi değerleri için dağılım grafikleri Şekil 2'de sunulmuştur.
- En iyi performansı sergileyen Model 9 incelendiğinde 4 veri grubundan da en az 2'ser değişken içermekte olduğunu gözlemlenmektedir. Dolayısıyla farklı veri gruplarından bazı değişkenleri birleştirmenin sonuçları önemli derece iyileştirdiği görülmüştür.
- Kullanılan yöntemden bağımsız olarak Model 11'den OUG ve Model 17'den SMO değişkenleri çıkarıldığında RMSE değerlerinin arttığı ve R^2 değerlerinin düştüğü görülmektedir. Bu nedenle bu değişkenlerin mutluluk düzeyi tahmininde önemli rol oynadığı gözlemlenmiştir.
- Kullanılan yöntemden bağımsız olarak Model 2'den ID, Model 7'den KTM ve Model 8'den OOE değişkenleri çıkarıldığında RMSE değerlerinin büyük oranda düştüğü ve R^2 değerlerinin arttığı gözlemlenmiştir. Bu nedenle mutluluk düzeyi tahmininde bu değişkenlerin olumsuz etkisi olduğu görülmektedir.
- Mutluluk düzeyi tahmininde kullanılan 3 yöntem karşılaştırıldığında tüm modeller için SVM yönteminin en düşük RMSE değerlerini ve en yüksek R^2 değerlerini verdiği, MLP tabanlı modellerin ise genelde TB tabanlı

¹ DTREG Predictive Modelling Software, <http://www.dtreg.com>

modellere göre daha düşük RMSE değerleri ve daha yüksek R^2 değerleri ürettiği gözlemlenmiştir. SVM tabanlı modellerin MLP ve TB tabanlı modellere göre RMSE değerlerini iyileştirme oranları Şekil 3'te görselleştirilmiştir.

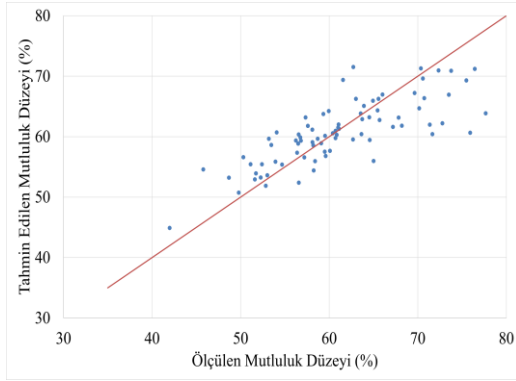
- En iyi yöntem olan SVM tabanlı modellerin ortalama RMSE değeri, MLP tabanlı modellerin ortalama RMSE

değerine göre %7.69, TB tabanlı modellerin ortalama RMSE değerine göre ise %8.88 daha düşüktür.

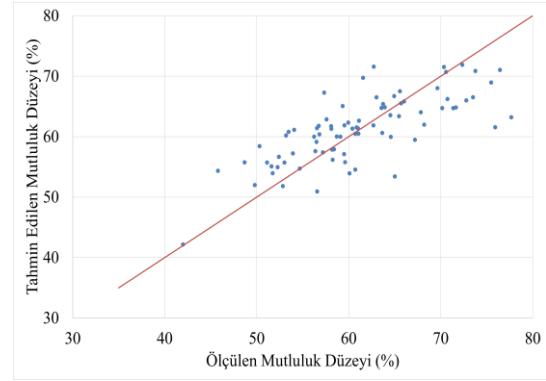
- Kullanılan yöntemler arasında en hızlı sonuçlar SVM tarafından elde edilmiştir. Eğitim süresi derlenen modele bağlı olarak SVM yöntemi için 1 ve 2 s arasında, MLP yöntemi için 3-5 s arasında ve TB yöntemi için 2-6 s arasında değişmektedir.

Tablo 6. SVM, MLP ve TB tabanlı tahmin modellerinin RMSE ve R^2 değerleri

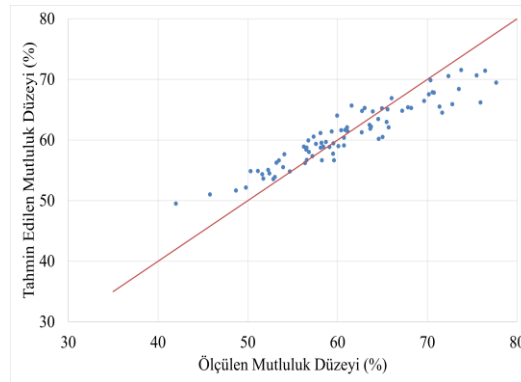
Model No	SVM		MLP		TB	
	RMSE	R^2	RMSE	R^2	RMSE	R^2
Model 1	5.55	0.44 (44.94%)	6.26	0.29 (29.98%)	6.17	0.32 (32.16%)
Model 2	6.09	0.33 (33.88%)	6.14	0.32 (32.67%)	6.16	0.32 (32.19%)
Model 3	5.79	0.40 (40.09%)	5.99	0.36 (36.07%)	6.02	0.35 (35.42%)
Model 4	5.46	0.46 (46.73%)	6.16	0.32 (32.32%)	6.24	0.30 (30.47%)
Model 5	5.43	0.47 (47.48%)	5.90	0.37 (37.79%)	5.92	0.37 (37.41%)
Model 6	5.45	0.47 (47.08%)	5.71	0.41 (41.81%)	5.91	0.37 (37.58%)
Model 7	5.54	0.45 (45.23%)	5.71	0.41 (41.83%)	5.88	0.38 (38.41%)
Model 8	5.46	0.46 (46.72%)	6.03	0.35 (35.13%)	6.17	0.31 (31.97%)
Model 9	5.32	0.49 (49.53%)	5.70	0.41 (41.97%)	5.84	0.39 (39.20%)
Model 10	5.37	0.48 (48.57%)	5.79	0.40 (40.11%)	5.91	0.37 (37.63%)
Model 11	5.40	0.47 (47.97%)	6.10	0.25 (25.16%)	6.12	0.33 (33.26%)
Model 12	5.50	0.45 (45.98%)	6.48	0.29 (29.62%)	6.54	0.23 (23.63%)
Model 13	5.60	0.43 (43.98%)	6.28	0.33 (33.01%)	6.29	0.29 (29.50%)
Model 14	5.64	0.43 (43.15%)	6.13	0.35 (35.64%)	6.14	0.32 (32.75%)
Model 15	5.85	0.38 (38.89%)	6.01	0.29 (29.50%)	6.13	0.32 (32.90%)
Model 16	5.64	0.43 (43.27%)	6.32	0.28 (28.83%)	6.33	0.28 (28.59%)
Model 17	5.78	0.40 (40.38%)	5.93	0.37 (37.28%)	6.12	0.33 (33.14%)
Model 18	5.96	0.36 (36.58%)	6.17	0.32 (32.15%)	6.48	0.25 (25.14%)



(a) SVM

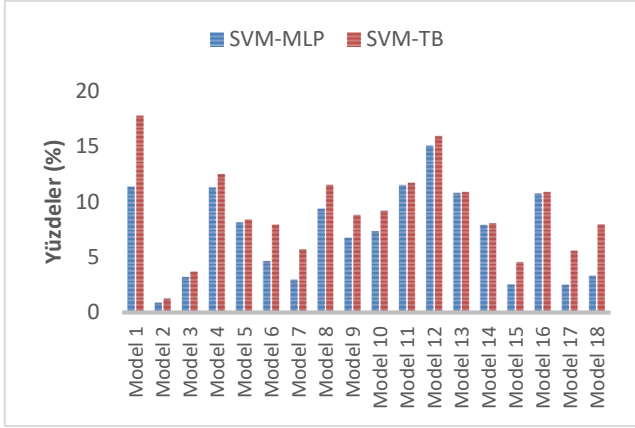


(b) MLP



(c) TB

Şekil 2. Eğitim aşamasında Model 9 için gerçek ve tahmin edilen mutluluk düzeyi değerleri için dağılım grafikleri



Şekil 3. SVM tabanlı modellerin MLP ve TB tabanlı modellere göre RMSE değerlerini iyileştirme oranları

5 Özet ve öneriler

Bu çalışma kapsamında Türkiye’deki mutluluk düzeyini etkileyen faktörleri belirlemek üzere Relief-F nitelik seçme algoritması ile birleştirilmiş Destek Vektör Makinesi (SVM), Çok Katmanlı Sinir Ağı (MLP) ve Ağaç Arttırma (TB) tabanlı yeni modeller önerilmiştir. Modellerin performansı RMSE ve R^2 değerlerinin hesaplanmasıyla değerlendirilmiştir. Elde edilen sonuçlara göre sağlık hizmetlerinden memnuniyet oranı, sağlığından memnuniyet oranı ile orta ve üstü gelir grubundaki hanelerin oranı değişkenlerinin mutluluk düzeyi tahmininde önemli rol oynadığı gözlemlenmiştir. Ayrıca, SVM yönteminin mutluluk düzeyi tahmininde etkin bir yöntem olduğu ve tüm tahmin modelleri için MLP ve TB yöntemlerine göre daha düşük RMSE ve daha yüksek R^2 değerleri ürettiği gözlemlenmiştir.

Mutluluk düzeyi tahmini ile ilgili gelecek çalışmalarda derin öğrenme gibi farklı makine öğrenme yöntemleri kullanılabilir ve nitelik seçici algoritmalar çeşitlendirilebilir. Daha fazla TÜİK veri grubu kullanılarak tahmin doğruluğu daha da artırılmış yeni modeller geliştirilebilir. Bu çalışmada önerilen en doğru tahmin modellerini içeren bir uygulama geliştirilerek kullanıcılara yeni verilerle ilgili gerçek zamanlı mutluluk düzeyi tahmini üretme imkânı sağlanabilir.

Teşekkür

Bu çalışmada, TÜİK’in “İllerde Yaşam Endeksi” ve “Yaşam Memnuniyeti Araştırması” kapsamında elde edilen veriler kullanılmıştır. Yazarlar veri setinin kullanım izni için TÜİK’e teşekkür eder.

Çıkar çatışması

Yazarlar çıkar çatışması olmadığını beyan etmektedir.

Benzerlik oranı (iThenticate): %17

Kaynaklar

- [1] N. Ahn, J. R. Garcia and J. F. Jimeno, The impact of unemployment on individual well-being in the EU. European Network of Economic Policy Research Institutes, Working paper, No: 29, 2004
- [2] Türkiye İstatistik Kurumu TÜİK 2016. Yaşam Memnuniyeti Araştırması Tarihçe, http://tuikweb.tuik.gov.tr/MicroVeri/YMA_2016/metaveri/tarihce/index.html, Erişim Tarihi 22.05.2020
- [3] R. Veenhoven, R. The study of life satisfaction. In W. E. Saris, R. Veenhoven, A. C. Scherpenzeel, & B. Bunting (Eds.), A comparative study of satisfaction with life in Europe, 11–48, 1996
- [4] Birleşmiş Milletler 2019 Yılı Dünya Mutluluk Raporu, https://en.wikipedia.org/wiki/World_Happiness_Report, Erişim Tarihi 05.09.2020
- [5] Türkiye İstatistik Kurumu TÜİK 2016. Yaşam Memnuniyeti Araştırması Veri Tabanı, http://tuikweb.tuik.gov.tr/MicroVeri/YMA_2016/yayinlar/index.html, Erişim Tarihi 05.09.2020
- [6] S. Selim, Life satisfaction and happiness in Turkey, Social Indicators Research, 88, 531-562, 2008. <https://doi.org/10.1007/s11205-007-9218-z>
- [7] G. M. Dyrddal, E. Roysamb, R. B. Nes, and J. Vitterso, Can a happy relationship predict a happy life? A population-based study of maternal well-being during the life transition of pregnancy, infancy, and toddlerhood, Journal of Happiness Studies, 12, 947-962, 2010. <https://doi.org/10.1007/s10902-010-9238-2>
- [8] I. Mirkhan, I. Shakerinia, M. Kafi, and N. Khalilzade, Prediction of life satisfaction based on emotional intelligence, happiness and religious attitude among female teachers of Urmia city, northwest of Iran. International Journal of School Health, 3, 1-5, 2014. <https://dx.doi.org/10.17795/intjsh-25144>
- [9] L. Noei Iran and M. Mahmoud Alilou, Predicting quality of life and happiness based on attachment style. Iranian Journal of Rehabilitation Research in Nursing, 4, 59-68, 2015
- [10] N. Jaques, S. Taylor, A. Azaria, A. Ghandeharioun, A. Sano, and R. Picard, Predicting students’ happiness from physiology, phone, mobility, and behavioral data. International Conference on Affective Computing and Intelligent Interaction, 21-24 Eylül, Xi’an China, 325-332, 2015
- [11] A. Caner, Happiness and life satisfaction in Turkey in recent years, Social Indicators Research, 127, 361-399, 2016. <https://doi.org/10.1007/s11205-015-0948-z>
- [12] M. Garaigordobil, Predictor variables of happiness and its connection with risk and protective factors for health. Frontiers in Psychology, 6, 1-10, 2015. <https://doi.org/10.3389/fpsyg.2015.01176>
- [13] İllerde Yaşam Endeksi, <http://www.resmiistatistik.gov.tr/detail/subject/illerde-yasam-endeksi/>, Erişim Tarihi: 05.09.2020
- [14] Türkiye İstatistik Kurumu - TÜİK, Yaşam Memnuniyeti Araştırması amaç, <https://tuikweb.tuik.gov.tr/>

gov.tr/MicroVeri/YMA_2016/metaveri/amac/index.html
Erişim Tarihi 05.09.2020

- [15] Türkiye İstatistik Kurumu - TÜİK Yaşam Memnuniyeti Araştırması, https://tuikweb.tuik.gov.tr/MicroVeri-YMA_2016/metaveri/kapsam/index.html, Erişim Tarihi: 05.09.2020
- [16] H. Yurtoğlu, Yapay sinir ağları metodolojisi ile öngörü modellenmesi: Bazı makroekonomik değişkenler için

Türkiye örneği, DPT - Uzmanlık Tezleri Yayın No: 2683, 2005

- [17] A. Pradhan, Support vector machine - A survey. International Journal of Emerging Technology and Advanced Engineering, 8, 82-85, 2012. <http://dx.doi.org/10.14257/ijta.2015.8.1.18>

