

Yapay Zekâ Algoritmaları ile Dünya Mutluluğunu Tahminleme

Abdulmuttalip DURAN¹, Kübra ÇELİK², Kemal ADEM^{3*} ve Sezer DÜMEN^{4*}

¹Elektrik Elektronik Mühendisliği /Savunma Teknolojileri, Sivas Bilim ve Teknoloji Üniversitesi, Türkiye

²Bitki Koruma / Tarım Bilimleri, Sivas Bilim ve Teknoloji Üniversitesi, Türkiye

³Bilgisayar Mühendisliği/ Sivas Bilim ve Teknoloji Üniversitesi, Türkiye

⁴Elektrik Elektronik Mühendisliği /Sivas Bilim ve Teknoloji Üniversitesi, Türkiye

*(sdumen@sivas.edu.tr)

Özet – Mutluluk bireyin yaşam kalitesini değerlendirme biçimidir ve ekonomik, sosyal, ulusal, kişisel ve coğrafi etmenlerden etkilenmektedir. Mutluluğu etkileyen bu faktörler makine öğrenmesi yöntemlerinde kullanılarak dünya ülkelerinin mutluluğu tahmin edilebilmektedir. Bu çalışmada; makine öğrenmesi algoritmaları kullanılarak dünya ülkelerinin mutluluğu tahmin edilmiştir. Makine öğrenmesi yöntemlerinden SVM, Ensemble, GPR, Tree, Linear Regression ve Neural Network algoritmaları kullanılmıştır. Dünya ülkelerinin mutluluğunu tahminlemede en başarılı makine öğrenmesi algoritması varsayılan hiperparametreler kullanıldığında 0.48281 RMSE değeriyle SVM algoritmasıdır. Daha sonra hiperparametre optimizasyonu yapılmıştır. Optimizasyon sonrasında ise en başarılı yöntem 0.45746 RMSE değeriyle yine SVM algoritmasıdır.

Anahtar Kelimeler – Mutluluk, Yapay Zeka, Makine Öğrenmesi, Regresyon, Destek Vektör Makineleri

I. GİRİŞ

Mutluluk, bireyin kendi yaşam kalitesini değerlendirdiği, uzun vadede hayattan duyduğu memnuniyet ya da psikolojik zenginliğe ulaşma arzusudur [1]. Başka bir deyişle, genel olarak bireyin yaşantısındaki olumlu duygu ve düşüncelerin olumsuz duygu ve düşüncelerden daha fazla olmasıdır [2].

İnsanlık tarihi kadar uzun bir geçmişe sahip olan mutluluk ilk kez 1970’lerde Bhutan tarafından sayısal bir ölçü olarak değerlendirilmiştir [3]. Mutluluk çok boyutlu bir kavram olduğundan dolayı analizi zor ve karmaşıktır. Bu zorluğun sebebi mutluluğun ülkeye, bölgeye göre farklılıklar gösterip farklı kesimler arasında etkilerinin değişkenlik göstermesi şeklinde açıklanabilir [4]. Bu etkiler; sosyal yaşantı, çalışma hayatı, toplumsal ilişkiler, başarı seviyesi ve verimlilik şeklinde sıralanabilir [1]. Mutluluğun önemi üzerine, Birleşmiş Milletler Sürdürülebilir Kalkınma Çözümleri Ağı 2012’den beri Dünya Mutluluk raporları yayınlamaktadır. Bu raporlarda: kişi başına düşen milli gelir (GSYİH), sosyal destek,

sağlıklı yaşam beklentisi, yolsuzluk algıları yer almakla birlikte sadece bunlarla sınırlı kalmayıp çeşitli faktörleri kullanarak mutluluğu açıklamaya çalışmaktadır [5]. Bu faktörler bir araya gelince mutluluğun hangi kriterlere göre tahmin edileceği önemli bir konu haline gelmiştir. Tahminlemenin istatistiksel hesaplamalar ile yapılması zor ve zaman aldığından dolayı günümüzde teknolojinin ilerlemesi ile birlikte makine öğrenmesi yöntemleri kullanılarak tahminleme işlemlerini yapmak çok daha hızlı, kolay ve kararlı sonuçlar verebilmektedir. Makina öğrenmesi yöntemleri ile Dünya mutluluk raporunu tahminleme üzerine yapılan çalışmalar çeşitli literatürler ile karşımıza çıkmaktadır. Literatür taramasında dünya mutluluk raporu ile ilgili 5 farklı çalışma incelenmiştir.

“Makine Öğrenimi Kullanarak Yaşam Kalitesini Tahminleme: Dünya Mutluluk İndeksi Vakası” adlı çalışmada 2015 – 2021 yılları arasındaki dünya mutluluk raporu veri setinden 137 kayıt 6 öznitelik kullanılmışlardır. 2015 – 2020 yılları arasındaki verileri eğitim için kullanmışlardır. 2021 yılı verisini test verisi olarak kullanmışlardır. Yaptıkları

çalışmada Lasso Regression, Multiple Linear Regression, Random Forest Regressor, Support Vector Regression, Gradient Boosting Regressor, XGboost Regressor, AdaBoost Regressor makine öğrenmesi tekniklerini kullanmışlardır. Analiz sonucunda en iyi performansı 0,8954 R-Squared Puanı ve 0,0565 RMSE değeri ile Lasso Regression da elde etmişlerdir [3].

“Makine Öğrenme Tekniklerini Kullanarak Dünya Mutluluğunu Anlama” adlı çalışmada 2019 Dünya mutluluk raporu veri setinden yola çıkılarak farklı makine öğrenme tekniklerini kullanmışlardır. 156 kayıt ve 7 öznitelik den oluşan veri seti üzerinde Decision Tables, Random Forest, SMOreg makine öğrenmesi yöntemleri kullanılarak Weka yazılım platformu aracılığıyla analizler yapmışlardır. Yapılan bu analizler sonucunda 0,0778 RMSE değeriyle en iyi Random Forest yönteminde başarı sağlamışlardır [4].

“Makine Öğrenme Yaklaşımlarını Kullanarak Dünya Mutluluk Raporu Veri Seti'nin Analizi” adlı çalışmada Kaggle'dan alınan 155 kayıt 7 öznitelikten oluşan veri seti üzerinde denetimli makine öğrenmesi teknikleri kullanılarak mutluluğu düşük, orta ve yüksek şeklinde puanlama yaparak sınıflandırma işlemi uygulamışlardır. NN, RF, XGB makine öğrenme metotlarını kullanarak veri seti üzerinde sınıflandırma analizleri yapmışlardır.

için 2016-2018 verileri kullanılıp performansları 2019 verileri üzerinden değerlendirme yapmışlardır. Yapılan bu analizler sonucunda, yaşam merdiveni tahminlerinin diğer tahminleme modellerine göre en iyi performansı MAE (0.29), MSE (0.15), R^2 (0.88) değerleriyle GRNN modelinde elde etmişlerdir [5].

“Kısa Ekolojik Anlık Ses Kayıtlarını Kullanarak Refahı Tahmin Etme” adlı çalışmada; 2019 yılı EMA anketinden alınan 1247 kayıt ve 12 öznitelikten oluşan veri seti üzerinde elde edilen kayıtlardan ses ve metin özelliklerinin tahminini farklı makina öğrenme teknikleri ile tespit etmişlerdir. Random Forest, SVM, MLP, Decision Tree, AdaBoost, Linear Regression ve Non-Linear Kernels makina öğrenme tekniklerini kullanmışlardır. Yapılan bu analizler sonucunda, en iyi tahminleme 0,665 RMSE ve 0,522 MAE değeriyle SVM algoritmasında tespit etmişlerdir [7].

Bu çalışmada; Kaggle'den alınan 2018 yılı dünya mutluluk raporu veri setine yeni coğrafi öznitelikler eklenilerek veri seti güncellenmiştir. Güncellenen veri seti üzerinde farklı makine öğrenme algoritmaları kullanılarak dünyadaki ülkelerin mutluluğu tahmin edilmiştir. Makine öğrenmesi algoritmaları kullanılarak dünyadaki ülkelerin mutluluğu tahmin edilmiştir

II. MATERYAL VE YÖNTEM

Over all rank	Country or region	Score	GDP per capita	Social support	Healthy life expectancy	Freedom to make life choices	Generosity	Perceptions of corruption
1	Finland	7.632	1.305	1.592	0.874	0.681	0.202	0.393
2	Norway	7.594	1.456	1.582	0.861	0.686	0.286	0.340
3	Denmark	7.555	1.351	1.590	0.868	0.683	0.284	0.408
4	Iceland	7.495	1.343	1.644	0.914	0.677	0.353	0.138
5	Switzerland	7.487	1.420	1.549	0.927	0.660	0.256	0.357

Yapılan analizler sonucunda en iyi sonuç %88 doğruluk değeriyle NN makine öğrenmesi metodu ile belirlenmiştir. Özniteliklerden, GSYİH'ın yaşam mutluluğu puanını etkileyen en önemli özellik olduğunu ve bunu sağlıklı yaşam beklentisinin izlediğini tespit etmişlerdir [6].

Dünya Mutluluğunu Araştırmak İçin Ağ Öğrenme Yaklaşımları” adlı araştırmada; Gallup dünya anketinden alınan 139 kayıt ve 12 öznitelikten oluşan veri seti üzerinde GRNN, Deep NN, Random Forest, XGBoost, OLS, Ridge, Decision Trees farklı tahminleme modellerinin analizini yapmışlardır. Tahminlemeye dayalı makine öğrenmesi modelleri

Bu çalışmada; Kaggle veri tabanından alınan 156 kayıt 6 öznitelikten oluşan 2018 yılı dünya mutluluk raporu veri seti üzerinde farklı öznitelikler eklenmesi ile güncellenmiştir. Ham veri setinin ilk beş satırı

Tablo 1. Dünya mutluluk raporu ham veri seti

'de gösterilmiştir.

Ham veri setinden yolsuzluk algısı özniteliği verisi eksik olduğu için Birleşik Arap Emirlikleri

ülkesi çıkarılmıştır ve güncel kayıt sayısı 155 olmuştur. Bu ham veri setine ülkelerin coğrafi

Overall rank	Country or region	Score	GDP per capita	Social support	Healthy life expectancy	Freedom to make life choices	Generosity	Perceptions of corruption
1	Finland	7.632	1.305	1.592	0.874	0.681	0.202	0.393
2	Norway	7.594	1.456	1.582	0.861	0.686	0.286	0.340
3	Denmark	7.555	1.351	1.590	0.868	0.683	0.284	0.408
4	Iceland	7.495	1.343	1.644	0.914	0.677	0.353	0.138
5	Switzerland	7.487	1.420	1.549	0.927	0.660	0.256	0.357

konumlarını belirten paralel ve meridyen numaraları eklenmiştir. Veri setinin daha anlaşılır olması için ülkelerin kıta bilgileri eklenmiştir.

Tablo 1. Dünya mutluluk raporu ham veri seti

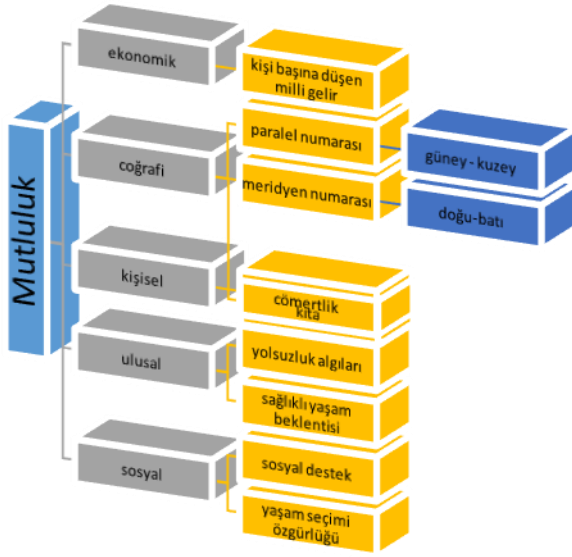
Tablo 2. Özniteliklerin standart sapma ve sınır değerleri

	Ortalama	Standart sapma	Minimum	Maksimum
Mutluluk puanı	5.366897	1.117433	2.905	7.632
Kişi başına düşen milli gelir	0.883677	0.380942	0.000	1.649
Sosyal destek	1.216058	0.301286	0.000	1.644
Sağlıklı yaşam beklentisi	0.596877	0.248312	0.000	1.030
Yaşam seçimi özgürlüğü	0.455606	0.162367	0.000	0.724
Cömertlik	0.180974	0.098789	0.000	0.598
Yolsuzluk algıları	0.112000	0.096492	0.000	0.457

Özniteliklerin ortalama, standart sapma, minimum ve maksimum değerleri tablo 2 de verilmiştir. Bir öz niteliğin ortalama değeri tüm değerlerin toplamının değer sayısına bölünmesi ile bulunmaktadır. Standart sapma değeri verilerin ortalamadan farklarının karelerinin toplamının veri sayısına bölümünün kareködür. Matematiksel ifadesi **Hata! Başvuru kaynağı bulunamadı.** de gösterilmiştir.

$$s = \sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})^2} \quad (1)$$

Eşitlik 1'de standart sapma hesaplaması gösterilmiştir. Burada s değişkeni standart sapmayı, N değişkeni toplam eleman sayısını, \bar{x} değeri ise aritmetik ortalamayı vermektedir.



Şekil 1.Öznitelik kategorileri

Minimum ve maksimum değerleri ise öz niteliklerin en düşük ve en yüksek değerlerini belirtmektedir. Mutluluğun öznitelikler üzerindeki etkisi; ekonomik, coğrafi, kişisel, ulusal ve sosyal olarak kategorileştirilmiştir. Özniteliklerden ekonomi kategorisi, kişi başına düşen milli gelir olarak ifade edilmiştir. Kişi başına düşen milli gelir (GSİYH) bir ülkenin gayri safi mili hasılasının o ülkenin nüfusuna bölünmesiyle elde edilir. Mutluluğu etkileyen faktörlerden coğrafi öznitelik; paralel, meridyen numarası ve ülkelerin hangi kıtalarda yer aldığına göre ülkelerin doğu-batı ve güney- kuzey konumuna göre kategorileştirilip konumun mutluluk üzerindeki etkisi incelenmiştir. Mutluluğu etkileyen özniteliklerden cömertlik o ülkedeki yaşayan insanların bağış yapma durumunu temsil etmektedir. Ulusal özniteliklerden olan yolsuzluk algısı, hükümette ve işletmelerde olan yolsuzluk durumunu açıklamaktadır. Sağlıklı yaşam beklentisi ise Dünya Sağlık Örgütü'nün açıklamasına göre sadece fiziki açıdan değil ruhsal ve sosyal yaşam üzerindeki etkilerinin mutluluk üzerindeki etkilerini açıklamaktadır [8]. Mutluluğu etkileyen faktörlerden sosyal destek, bir kişinin zor zamanlarında güvenilebileceği birinin olup olmadığını açıklayan terim olarak ele alınmıştır. Yaşam seçimi özgürlüğü ise bir kişinin yaşamında ne yapacağını seçme özgürlüğünden memnun olup olmadığını soran yanıtın açıklamasıdır.

A. Yöntem

Veri seti hazırlandıktan sonra verileri analiz etmek için makine öğrenmesi algoritmaları çalıştırılmıştır. Öznitelikler üzerinde makine öğrenmesi algoritmaları kullanılarak ülkelerin mutluluk değerleri tahmin edilmiştir. Makine öğrenmesi yöntemleri ile tahminleme işlemi yapılırken veri setinin %80'i eğitim, %20'si ise test için kullanılmıştır. En başarılı makine öğrenmesi algoritmaları; DVM, GSR ve Topluluk algoritmalarıdır.

1) Destek Vektör Makinesi

Destek Vektör Makineleri (DVM) ilk olarak istatistiksel öğrenme teorisine dayalı bir sınıflandırma algoritması olarak ortaya çıkmıştır (Chapelle et al., 1999, Corinna & Vapnik, 1995). Daha sonraki yapılan çalışmalarda, alternatif bir kayıp fonksiyonunun başlatılmasıyla tahminleme analizleri için de kullanılabilirliği tespit edilmiştir [11]. Veri kümesinde birbiriyle benzerlik gösteren gruplar arasına birbirinden en uzak olan noktalardan sınırlar çizerek sınıflandırma veya tahminleme işlemleri gerçekleştiren makine öğrenmesi yöntemidir [12]. DVM algoritmasının performansı C değeri, çekirdek fonksiyonun tipi ve ilgili parametreleri ve ϵ değeri gibi çeşitli parametrelere bağlıdır.

2) Gauss Süreci Regresyonu

Gauss süreci regresyon (GSR) modelleri doğrusal olmayan problemleri çözmek amacıyla parametrik olmayan çekirdek tabanlı olasılık teknikleridir. GSR küçük verilerde dahi yüksek başarılı sonuçlar üretebilir ve tahminlerde belirsizlik ölçümleri yapabilecek niteliklere sahiptir. Gauss süreçleri istatistiksel modellemede, tahminleme analizlerinde ve haritalamanın daha yüksek boyutlarda analizinde kullanılır [13].

3) Ensemble Algoritması (Topluluk Yöntemleri)

Topluluk yöntemleri bir dizi sınıflandırıcı oluşturan ve ardından yeni veri noktalarının tahminlerini sınıflandıran öğrenme algoritmalarıdır. Orijinal topluluk yöntemi Bayes ortalaması almaz fakat yeni algoritmalar hata düzelten çıktı kodlamaları, paketleme ve artırma içerir [14]. Topluluk öğrenmesi öncelikle bir modelin performansını arttırmak veya kötü bir modelin

talihsiz bir şekilde seçilme olasılığını azaltmak için kullanılır [12].

Makine öğrenmesi algoritmalarının varsayılan hiper parametreleri ile tahminleme işlemi gerçekleştirildikten sonra, tahminleme sonucunun iyileştirilmesi, hata payının düşürülmesi için hiper parametre optimizasyonu Matlab üzerinden optimize algoritmaları kullanılarak gerçekleştirilmiştir. Matlab içerisindeki optimizasyon algoritmaları şunlardır.

Grid Search: Bu algoritma en temel ve en yavaş çalışan hiper parametre tespit algoritmasıdır. Bu teknikte sağlanan tüm hiper parametre değerleri teker teker denenerek en iyi sonucu veren algoritma seçilir. Tüm parametreleri denendiği için çok yavaş çalışır ama en uygun tespiti yapar [15].

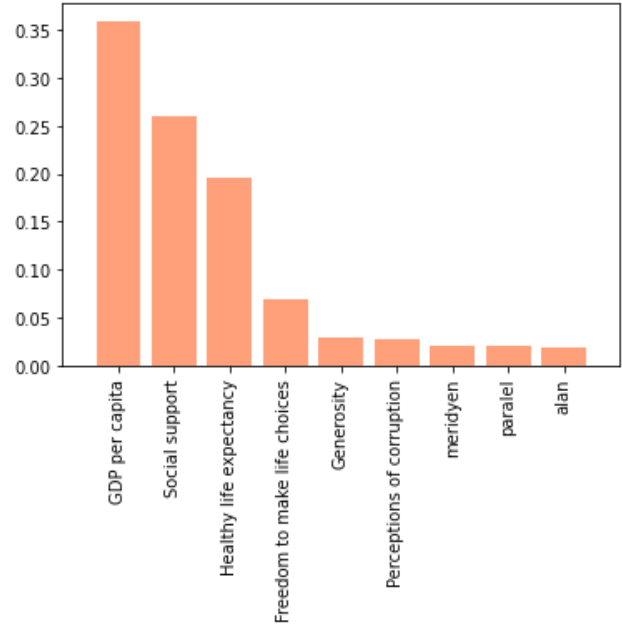
Random Search: Bu algoritma da Grid Search algoritmasına benzerdir fakat tüm parametreleri sırayla denemek yerine rasgele olarak parametreler seçerek dener. Parametrelerin tamamını denemediği için oldukça hızlı çalışır fakat en iyi parametreleri kaçırabileceği için Grid Search kadar doğru tespit yapamaz [15].

Bayesian: Doğruluk ve zaman tasarrufu açısından en doğru algoritmadır. Rastgele veya sırayla parametre aramak yerine hiper parametreleri geçmiş değerleri kullanarak oluşturduğu olasılık fonksiyonunu kullanarak tespit eder. Bu sayede bir sonraki parametre setini rastgele seçmek yerine, algoritma seçimini optimize eder ve en iyi parametre setini en kısa sürede tespit eder [15].

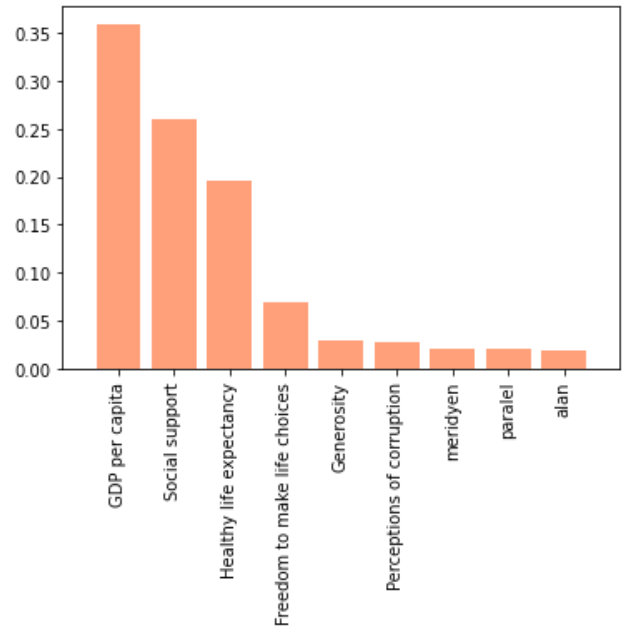
Bu projede Bayesian optimize yöntemini kullanarak hata değerinin en düşük olduğu hiper parametreyi tespit edilmiştir.

III. BULGULAR VE TARTIŞMA

Bu çalışma İ5 işlemcili, 8GB Ram'e sahip bir bilgisayarda Matlab programı kullanılarak gerçekleştirilmiştir. Özniteliklerin sonuca etkisi



'de gösterilmiştir.



Şekil 2. Öznitelik korelasyon dağılımı

Korelasyon dağılım tablosu özniteliklerin sonuç ile olan ilişkisini vermektedir. Şekil 2'de görüldüğü gibi ver setinin korelasyon dağılımına bakıldığında mutluluğu etkileyen en önemli faktörün kişi başına düşen milli gelir olduğu tespit edilmiştir. Diğer önemli iki öznitelik ise sosyal destek ve sağlıklı yaşam belirtisidir. Mutluluğu en az etkileyen özneliğin ise ülkelerin yüzey alanı olduğu tespit edilmiştir. Bu bilgiye göre ülkelerin mutluluk değerinin yüzey alanından çok etkilenmediği sonucu çıkmaktadır.

Bu değerlerin performansını değerlendirmek için çeşitli ölçütler kullanılmıştır. Bu ölçütlerden MAE değeri, Ortalama mutlak hata anlamına gelen, gerçek ve tahmin edilen değerler arasındaki mutlak farkın ortalamasının ölçen hataların ölçüsüdür [3]. MSE değeri, Ortalama kare hatası, tahmin edilen ve gerçek değerler arasındaki uyumsuzluğun ortalama kare hatasını ölçen bir kriterdir [3]. RMSE değeri, ortalama karekök hatası, ortalama kare hatasının kareköküdür. R-kare değeri ise, modelin verilerle nasıl eşleştiğini test etmek için kullanılan bir belirleme katsayısıdır. Model tarafından açıklanan bağımlı değişkenin değişkenlik ölçüsünü ölçer. Genel olarak girdi ve çıktı arasındaki doğrusal ilişkinin gücünü temsil eden bir korelasyon ölçüsüdür [3].

$$MAE = \frac{1}{m} \sum_{i=1}^m |y_i - y'_i| \quad (2)$$

$$MSE = \frac{1}{m} \sum_{i=1}^m (y_i - y'_i)^2 \quad (3)$$

$$RMSE = \sqrt{MSE} \quad (4)$$

$$R^2 = 1 - \frac{\sum_{i=1}^m (y_i - y'_i)^2}{\sum_{i=1}^m (y_i - \bar{y})^2} \quad (5)$$

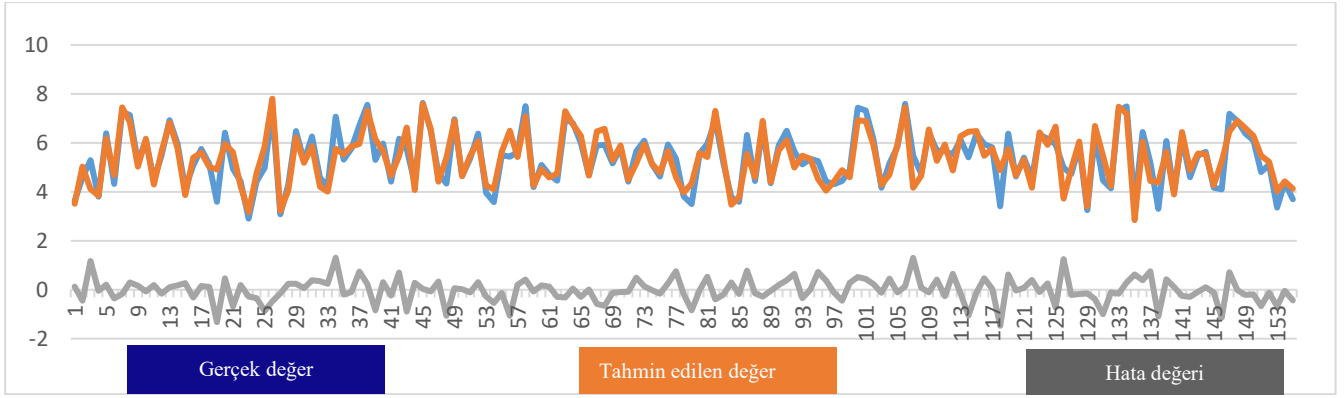
Eşitlik 2,3 ve 5 de kullanılan m değeri test verilerinin toplam eleman sayısını, y'_i değişkeni tahmin edilen değeri, y_i değeri ise gerçek değeri göstermektedir.

Projede data setine ön işlemler yapıldıktan ve yeni öznitelikler eklendikten sonra varsayılan hiperparametreler ile makine öğrenmesi algoritmaları çalıştırılmıştır. Varsayılan hiperparametre olarak Matlab da auto modu seçilmiştir. Ülke yüzey alanının tahminleme hatasını yükselttiği gözlemlenmiştir. Yüzey alanı özneliği veri setinden çıkartılarak tahminleme işlemlerine devam edilmiştir. Sonuçlar tablo 4’de gösterilmiştir.

Ayrıca, varsayılan hiper parametreler gerçek değer, tahmin edilen değer ve hatalar tek bir grafikte birleştirilerek **Hata! Başvuru kaynağı bulunamadı.**’te gösterilmiştir.

Tablo 3. Varsayılan hiper parametrelere göre çıktılar

Algoritmalar	Performans				
	RMSE	R ²	MSE	MAE	Eğitim Zamanı
Quadratic SVM	0.48281	0.81	0.2331	0.35885	3.7338 sec
Exponential GPR	0.49358	0.80	0.24362	0.36316	16.808 sec
Matern 5/2 GPR	0.49823	0.80	0.24824	0.36388	17.935 sec
Coarse Gaussian SVM	0.49831	0.80	0.24831	0.37026	6.5188 sec
Rational Quadratic GPR	0.5025	0.80	0.2525	0.3675	20.453 sec
Ensemble Bagged Trees	0.50692	0.79	0.25697	0.39726	17.56 sec
Stepwise Linear	0.51576	0.79	0.26601	0.37929	43.373 sec
Medium Gaussian SVM	0.51676	0.79	0.26704	0.37957	5.4977 sec



Şekil 3. Optimize öncesi gerçek değer ile tahmin edilen değer grafiği

Data setine eklenen coğrafi özniteliklerin mutluluğa etkisi şekil'3 ve şekil'4 de gösterilmiştir.

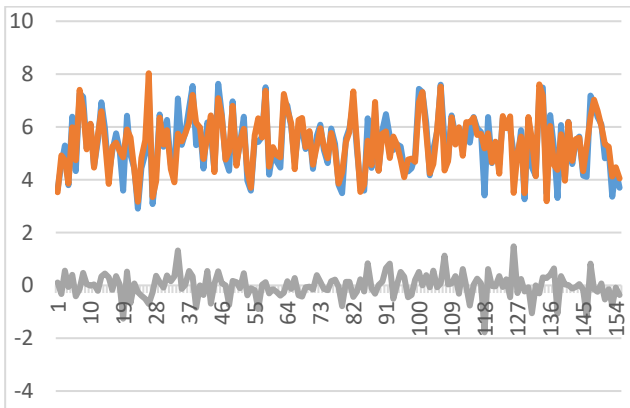
Matlab üzerinden optimize uygulaması kullanılarak makine öğrenmesi algoritmalarının en uygun hiper parametreleri tespit edilmiştir. Optimizasyon sonucu SVM algoritması için hiper parametre olarak Kernel Function değeri Cubic, Kernel Scale değeri 2.89, Box Constraint değeri 0.14369, Epsilon değeri 0.051651, Standardize Data değeri yes olarak seçilmiştir. Decision Tree algoritması için Minimum Leaf Size değeri 3 seçilmiştir. GPR algoritması için Basis Function

değeri Zero, Kernel Function değeri Nonisotropic Matern 3/2, Kernel Scale değeri 5.5661, Sigma değeri 0.00010013, Standardize Data değeri ise Yes olarak seçilmiştir. Ensemble algoritması için Ensemble Method değeri Bag, Minimum Leaf Size değeri 1, Number Of Learners değeri 91, Number Of Predictors To Sample değeri 8 olarak seçilmiştir. Neural Network algoritması için Number Of Fully Connected Layers değeri 1, Activation değeri Sigmoid, Regularization Strength (Lambda) değeri 0.00067692, Standardize Date değeri No, First Layer Size değeri ise 2 olarak seçilmiştir. Yeni sonuçlar tablo 5'te gösterilmiştir.

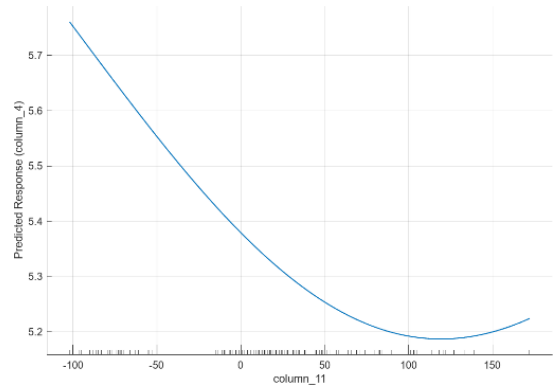
Tablo 4. Optimize hiperparametrelere göre çıktılar

Algoritmalar	Performans				
	RMSE	R ²	MSE	MAE	Eğitim Zamanı
SVM	0.45746	0.84	0.20927	0.33211	1.9359 sec
Ensemble	0.47619	0.82	0.22676	0.37235	8.9379 sec
GPR	0.48818	0.81	0.23832	0.35977	5.7719 sec
Tree	0.56256	0.75648	0.31648	0.42604	6.2947 sec
Neural Network	0.64216	0.67	0.41237	0.45463	18.749 sec

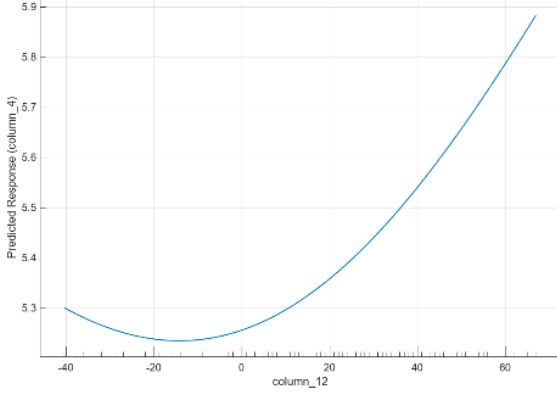
Gerçek değer, tahmin edilen değer ve hatalar tek bir grafikte birleştirilerek Şekil 4'te gösterilmiştir



Şekil 4. Gerçek değer ile tahmin edilen değer grafiği



Şekil 5. Ülkelerin meridyen numarasının mutluluğa etkisi



Şekil 6. Ülkelerin paralel numarasının mutluluğa etkisi

Şekillerde görüldüğü gibi ülkelerin paralel ve meridyen numaralarının mutluluk üzerinde etkisi bulunmaktadır. Meridyen numarası ülkelerin Greenwich bölgesine doğu veya batı tarafından uzaklığını göstermektedir. Meridyen grafiği incelendiğinde doğuya doğru gidildikçe ülkelerin mutluluğunun düştüğü gözlemlenmektedir. Benzer şekilde paralel numarası da ülkelerin ekvatora kuzey veya güney yönünden uzaklığını göstermektedir. Paralel grafiği incelendiğinde kuzeye doğru gidildikçe mutluluğun arttığı tespit edilmiştir.

IV. SONUÇLAR

İnsanlık için önemli bir yere sahip olan mutluluk, ekonomik, kişisel, ulusal vb. birçok etmenlerden etkilenmektedir. Mutluluğu etkileyen etmenler makine öğrenmesi algoritmaları kullanılarak regresyon analizleri yapılmıştır. Literatürde bu konuda regresyon ile ilgili çok sayıda makine öğrenmesi çalışmaları görülmüştür. Bu çalışmada Kaggle'dan alınan 2018 yılı dünya mutluluk raporu veri seti üzerine farklı öznitelikler eklenerek dünyadaki ülkelerin mutluluğu tahmin edilmiştir. Özniteliklerin korelasyon dağılımına göre mutluluğu etkileyen en önemli öznitelik "Kişi başına düşen milli gelir" olduğu tespit edilmiştir. Makine öğrenmesi algoritmalarından en başarılı yöntem 0.458 RMSE değeriyle SVM algoritmasının diğer yöntemlere göre daha iyi performans sergilediği görülmüştür. Bu değere hiperparametre optimizasyonu uygulandıktan sonra hiperparametreler; Kernel Function değeri Cubic, Kernel Scale değeri 2.89, Box Constraint değeri 0.14369, Epsilon değeri 0.051651, Standardize Data değeri yes olarak tespit edilmiştir.

KAYNAKLAR

- [1] İ. Yıldız and F. Abut, "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," vol. 11, no. 1, pp. 1–9, 2022, doi: 10.28948/ngmuh.873199.
- [2] G. Yazici Çelebi, "BİREYLERİN MutluluDüzeylerininDeğişkenler Açısındaİncelenmesi Examination of Happiness Levels of Individuals in Terms of Various Variables," vol. 3, no. 1, pp. 179–189, 2020.
- [3] A. Jannani, N. Sael, and F. Benabbou, "Predicting Quality of Life using Machine Learning: case of World Happiness Index," 2021 4th Int. Symp. Adv. Electr. Commun. Technol. ISAECT 2021, 2021, doi: 10.1109/ISAECT53699.2021.9668429.
- [4] F. Ibrat, J. Gyalmo, Z. Alom, M. A. Awal, and M. A. Azim, "Understanding World Happiness using Machine Learning Techniques," 6th Int. Conf. Comput. Commun. Chem. Mater. Electron. Eng. IC4ME2 2021, no. December, pp. 26–27, 2021, doi: 10.1109/IC4ME253898.2021.9768407.
- [5] Özkaya, U., Öztürk, Ş., & Barstugan, M. (2020). Coronavirus (COVID-19) classification using deep features fusion and ranking technique. Big Data Analytics and Artificial Intelligence Against COVID-19: Innovation Vision and Approach, 281-295.
- [6] M. A. Khder, M. A. Sayfi, and S. W. Fujo, "Analysis of World Happiness Report Dataset Using Machine Learning Approaches," Int. J. Adv. Soft Comput. its Appl., vol. 14, no. 1, pp. 14–34, 2022, doi: 10.15849/IJASCA.220328.02.
- [7] Y. N. Huang, S. Zhao, M. L. Rivera, J. I. Hong, and R. E. Kraut, "Predicting Well-being Using Short Ecological Momentary Audio Recordings," Conf. Hum. Factors Comput. Syst. - Proc., 2021, doi: 10.1145/3411763.3451629.
- [8] Ö. U. Ğ. U. Ğ. Lu and Y. Ç. E. L. İ. K, "Sa ğ l ı k Sistemleri Performans Ölçümü , Önemi ve Dünya Sa ğ l ı k Örgütü Yakla ş ı m ı," 2000.
- [9] O. Chapelle, P. Haffner, and V. N. Vapnik, "Support vector machines for histogram-based image classification," IEEE Trans. Neural Networks, vol. 10, no. 5, pp. 1055–1064, 1999, doi: 10.1109/72.788646.
- [10] C. Corinna and V. Vapnik, "Support-Vector Networks," AT&T Bell Labs., Hohndel, NJ 07733, USA, vol. 628, no. 1, pp. 273–297, 1995, doi: 10.1088/1742-6596/628/1/012073.
- [11] T. B. Trafalis and H. Ince, "Support vector machine

- for regression and applications to financial forecasting,” *Proc. Int. Jt. Conf. Neural Networks*, vol. 6, no. x, pp. 348–353, 2000, doi: 10.1109/ijcnn.2000.859420.
- [12] U. Jain, K. Nathani, N. Ruban, A. N. J. Raj, Z. Zhuang, and V. G. V. Mahesh, “Cubic SVM classifier based feature extraction and emotion detection from speech signals,” *Proc. - 2018 Int. Conf. Sens. Networks Signal Process. SNSP 2018*, pp. 386–391, 2019, doi: 10.1109/SNSP.2018.00081.
- [13] D. Yergök and M. Acı, “Toplu Yemek Üretiminde Günlük Talep Tahmini için Alternatif Bir Yaklaşım: Öğrenci Regresyon,” *Eur. J. Sci. Technol.*, no. October, pp. 64–73, 2019, doi: 10.31590/ejosat.636368.
- [14] T. G. Dietterich and Oregon, “Ensemble methods in machine learning. In: International Workshop on Multiple Classifier Models,” *Oncogene*, vol. 12, no. 2, pp. 1–15, 2000.
- [15] H. Alibrahim and S. A. Ludwig, “Hyperparameter Optimization: Comparing Genetic Algorithm against Grid Search and Bayesian Optimization,” *2021 IEEE Congr. Evol. Comput. CEC 2021 - Proc.*, pp. 1551–1559, 2021, doi: 10.1109/CEC45853.2021.9504761.