

El Yazısı ile Cinsiyet Tahmini

Metin KADIK ve Ezgi KAYA

Bilgisayar Mühendisliği
Kocaeli Üniversitesi

{150201166& 150201158 } @kocaeli.edu.tr

Bu çalışmada makine öğrenmesi yöntemleriyle, el yazıları üzerinden cinsiyet tespit etmek amaçlanmıştır. Makine öğrenmelerinin karşılaştırılması ve karmaşık veriler karşısındaki davranışları ve analizi çalışmanın içeriğini oluşturmaktadır.

Anahtar Kelimeler – El Yazısı, Cinsiyet Tanıma, , SVM, Lojistik Regresyon, Random Forest .

I.GİRİŞ

Günümüz dünyasında bilgi miktarı gün geçtikçe artmaktadır. Artan bilgiye kısa sürede erişmek için birbiri ile ilişkili bilgileri bulup onları sınıflandırmayı çalışmak gerekir. Sınıflandırmada ki amaç önceden toplana bilgilerin hangi sınıfa ait olduğu bilindiğinde gelecek olan bilgilerinde tahminsel bir modelle bulmaktır. Elimizdeki bilgilerle bir eğitim veri seti öğrenme gerçekleştikten sonra yeni gelen bilgilerle test veri seti kurularak modelin performansı ölçülür.

Bu çalışmamızın amacı cinsiyet etiketleri bulunan bir el yazısı verisinin eğitilmesi ve yüksek doğrulukta tahmin değerlerine ulaşılmasıdır.

Eğitim için ICDAR 2013 veri seti kullanılmıştır. İçinde Arapça ve İngilizce el yazıları bulunmaktadır. Bir çok ödüllü yarışmada kullanılan bir veri setidir[21].

Katar Üniversitesinin "Kaggle" üzerinden 2013 yılında düzenlediği bir yarışmada cinsiyet etiketleri yayınlanmıştır. Bu çalışmada, bu cinsiyet etiketlerine göre çeşitli denetimli öğrenme yöntemleri kullanılmıştır.

Denetimli öğrenme yöntemlerinden üç tanesi bu çalışma için uygun görülmüştür. Bunlar;

Rastgele Orman Algoritması , Destek Vektör Makineleri ve Lojistik Regresyon'dur.

Bu etiketli öğrenme yöntemleri özet ve sonuçlar kısmında detaylı olarak karşılaştırılacaktır.——

Veri setimizde yedi bin altmış altı (7066) adet özellik mevcuttur. Bu kadar çok özelliğin kullanımı gereksizdir. Sonucumuza en az etki eden, alakasız özellikler çıkarılmalıdır. En çok etki eden özelliklerin seçilmesi durumunda isabet oranının yükselmesi beklenmelidir[19].

Özellik seçiminin ardından veri modellere girdi olarak verilir. Optimum parametrelere sahip olan modellerin doğru tahmin yapabilme olasılığı artar.

II. ÖN İŞLEMLER

Veri setlerinde çok sayıda özellik olabilir. Bu özelliklerin elenmesi ve işe yarayan özelliklerin ortaya çıkartılması gereklidir. Bu işlem iyi bir makine öğrenmesi probleminin temel işlemlerindendir.

Bu problemde öncelikle özellikler arası korelasyona bakıldı. Korelasyon bize veriler arasındaki pozitif ve negatif uyumu göstermede önemli rol oynayan değerlerdendir[18].

Korelasyon katsayısı negatif ise iki değişken arasında ters ilişki vardır, yani "değişkenlerden biri artarken diğeri azalmaktadır" denir. Korelasyon katsayısı pozitif ise "değişkenlerden biri artarken diğerde artmaktadır" yorumu yapılır.

Verilerin normal dağılıma sahip olması durumunda Pearson korelasyon katsayısı, verilerin normal dağılmadığı durumda ise Spearman Rank korelasyon katsayısı tercih edilir[20].

Korelasyon katsayısı (r) nın yorumu;

$r < 0.2$ ise çok zayıf ilişki yada korelasyon yok

0.2-0.4 arasında ise zayıf korelasyon

0.4-0.6 arasında ise orta şiddette korelasyon

0.6-0.8 arasında ise yüksek korelasyon

$0.8 >$ ise çok yüksek korelasyon olduğu yorumu yapılır[18].

Yüksek korelasyona sahip iki özellikten birini eleyebiliriz. Çünkü birinden diğerrinin sonucu yaklaşık olarak bilinebilir. Bu eleme işlemi ile özellik sayısı düşürülür gereksiz özellikler atılmış olur. Ancak verideki özellik sayısı çok yüksek olması sebebiyle bu yöntemden vazgeçildi.

III. YÖNTEMLER

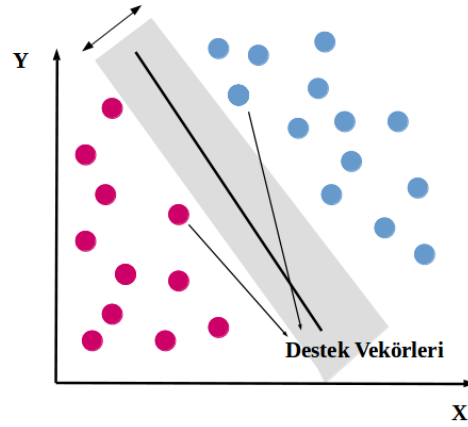
Bu bölümde makalede önerilen modelin arkasındaki kavramları anlamak için gerekli temel bilgileri sunacağız.

a. Destek Vektör Makineleri :

1963 yılında Vladimir Vapnik ve Alexey Chervonenkis tarafından temelleri atılmıştır. Her ne kadar temelleri 60'lı yıllara dayansa da 1995 yılında Vladimir Vapnik, Bernhard Boser ve Isabelle Guyon tarafından geliştirilmiştir[9]

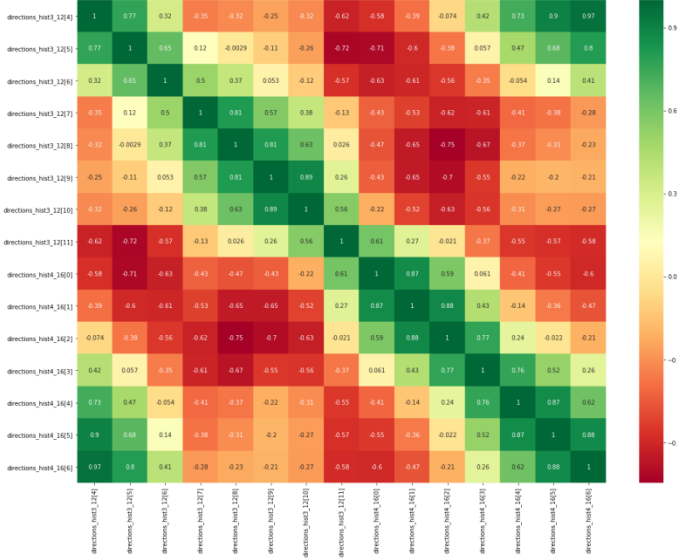
Destek Vektör Makineleri regresyon ve sınıflandırma problemleri için kullanılan bir denetimli makine öğrenmesi algoritmasıdır. [17] DVM regresyon metodundaki temel fikir, eldeki eğitim verilerinin karakterini mümkün olduğunca gerçeğe yakın bir şekilde yansıtan ve istatistiksel öğrenme teorisine uyan doğrusal ayırıcı fonksiyonun bulunmasıdır. Sınıflandırmaya benzer bir şekilde regresyonda da doğrusal olmayan durumların işlenebilmesi için çekirdek fonksiyonları kullanılır.

Destek Vektör Makineleri sınıflandırma konusunda kullanılan çok basit ve etkili bir yöntemdir. DVM'de amaç, sınıfları birbirinden ayıracak optimal hiper düzleminin elde edilmesidir. Başka bir ifadeyle farklı sınıflara ait destek vektörleri arasındaki uzaklığı maksimize etmektir.



Şekil 3. Destek Vektör Makinesi

Destek vektör makineleri veri setinin doğrusal ayrılma ve ayrılma durumlarına göre ikiye ayrılır.

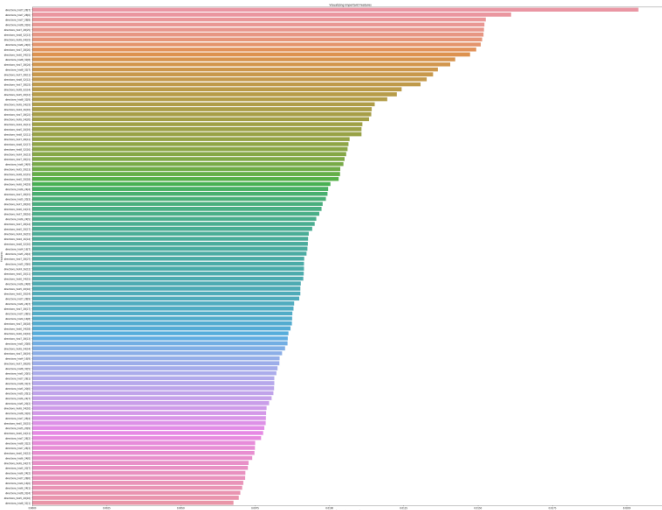


Şekil 1. Korelasyon matrisi-ısı haritası

İkinci olarak baktığımız teknik ise 'özellik önemi' tekniğidir.

Özellik önemi (Feature importance), verilerinizin her özelliği için bir puan verir, puan daha önemli veya alakalı ise çıktı değişkeninize yönelik özellik o kadar yüksek olur. Özellik önemi tekniği ağaç tabanlı algoritmaları ile entegre çalışır. Bu verisetinde özelliklere önem puanları verildi ve büyüktен küçüğe doğru sıralandılar. Özellik değeri 0.0002 den büyük olan özellikleri seçtik. Bu teknik ile 7066 olan özellik sayımızı 2019 a düşürmüştük.

Bu durum hem işlemci gücümüzü fazla kullanmamamızı sağladı hem de daha iyi bir isabetle tahmin yapabilmemizi.

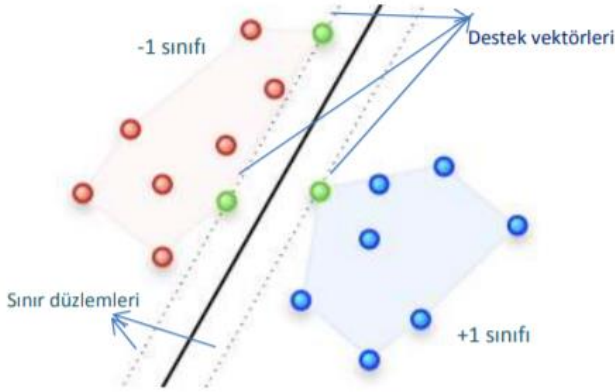


Şekil 2. Kısıtlı sayıda özelliğin önem puanına göre sıralanması

1-Doğrusal Destek Vektör Makineleri

Doğrusal Destek Vektör Makineleriyle sınıflandırmada, iki sınıfa ait örneklerin doğrusal olarak dağılmasıdır. Bu durumda bu iki sınıfın, eğitim verisi kullanılarak elde edilen bir karar fonksiyonu yardımıyla birbirinden ayrılmasıdır.

Veri setini ikiye ayıran doğruya karar doğrusu denir. Amaç optimal karar doğrusunu belirlemektir. Sınır çizgisinin, iki sınıfın sınır çizgilerine en yakın uzaklıkta olması gerekmektedir. Bu karar noktasına en yakın noktalar, destek noktaları olarak adlandırılmaktadır.



Şekil 4. Doğrusal Destek Vektör Makinesi

Destek vektör makineleri sınıflandırmada genellikle (-1,+1) şeklinde sınıf etiketleri kullanılmaktadır. [1] Kesik çizgi ile gösterilmiş düzlemlere sınır düzlemleri denir. Sınır düzlemlerinin tam ortasından geçen ve her iki düzleme de eşit uzaklıkta bulunan düzlem ise hiper düzlem olarak ifade edilir.

İkili sınıf verileri birbirinden ayırabilen birçok hiper düzlem çizilebilir. Yukarıda da belirttiğimiz gibi amaç kendisine en yakın noktalar arasındaki uzaklığı maksimuma çıkaran hiper düzlemi bulmaktır.

$$w \cdot x + b \quad (1)$$

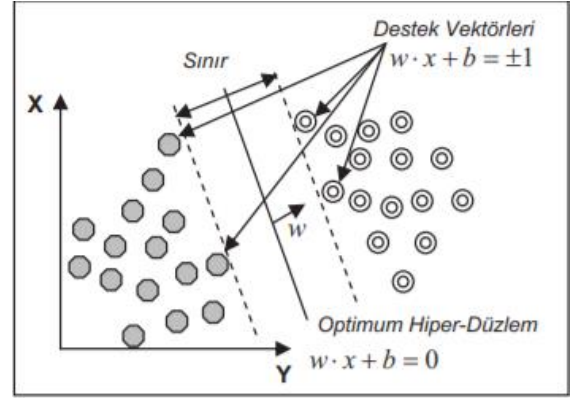
Doğrusal olarak ayrılabilen iki sınıflı bir sınıflandırma probleminde DVM'nin eğitimi için k adet veriden oluşan eğitim setinin $\{x_i, y_i\}$, $i = 1, 2, \dots, k$ olduğu kabul edilirse, hiper düzleme ait eşitsizlikleri sağlaması gerekir. [6]

$$w \cdot x_i + b_i \geq +1, y = +1 \quad (2)$$

$$w \cdot x_i + b_i \leq -1, y = -1 \quad (3)$$

Her x_i örneği için $x \in R^N$, N-Boyutlu girdi vektörü olup $y \in \{-1, +1\}$ örneklerin ait olduğu sınıf etiketlerini, w hiper düzlemin normali yani ağırlık vektörünü ve b eğilim değerini göstermektedir. [3]

Optimum hiper düzlemin bulunması için bu düzleme paralel ve sınırları oluşturacak iki hiper düzlem belirlenir. Bu düzlemi oluşturan noktalara yukarıda da bahsettiğimiz gibi destek vektörleri denilir. Bu düzlemler $w \cdot x_i + b_i = \mp 1$ şeklinde gösterilir.



Şekil 5. DDVM Optimum Hiper Düzlem

Optimum hiper düzlem sınırı maksimum yapan ve sınırın ortasından geçen düzlemdir. Optimum hiper-düzlem sınırının maksimuma çıkarılması için $\|w\|$ ifadesinin minimum hale getirilmesi gerekir. Burada $\|w\|$, w normal düzleminin normudur. En iyi hiper düzlemin belirlenmesi aşağıdaki sınırlı optimizasyon probleminin çözümünü gerektirir. [5]

$$\min \left[\frac{1}{2} \|w\|^2 \right] \quad (4)$$

$$\forall_i \text{ için } y_i ((w \cdot x_i) + b) - 1 \geq 0 \quad (5)$$

Optimizasyon probleminin çözülmesinde Lagrange çarpanlarından yararlanılır. Bu çarpanlar problemin daha kolay çözülmesine imkan verir. [7]

$$L(w, b, \alpha) = \frac{1}{2} \|w\|^2 - \sum_{i=1}^k \alpha_i y_i (w \cdot x_i + b) + \sum_{i=1}^k \alpha_i \quad (6)$$

Bu eşitlikte $\alpha_i \geq 0$ olmak üzere, her bir α_i Lagrange çarpanı olarak ifade edilir. Doğrusal ayrılabilen ikili sınıf problemi için karar fonksiyonu

$$f(x) = \text{sign}\left(\sum_{i=1}^k \alpha_i y_i(x \cdot x_i) + b\right) \quad (7)$$

şekilde ifade edilir.[3]

$$f(x) = h(\langle w^*, x_i \rangle + b^*) \Rightarrow \quad (8)$$

$$h(z) = \begin{cases} -1 & z < -1 \\ z & -1 \leq z \leq +1 \\ +1 & z > +1 \end{cases}$$

2- Doğrusal Olmayan Destek Vektör Makineleri

Veri setinin gürültülü veri içermesi, çok boyutlu olması veya karmaşık yapısıdır[Sevgi]. Bu durumda eğitim verilerinin bir kısmı optimum hiper düzlemin diğer tarafında kalmasından kaynaklanan problem pozitif bir yapay değişkenin (ξ_i) modele eklenmesi ile çözülür.[1]

$$f(x_i) = \langle w, x_i \rangle + b_i \geq +1 - \xi_i, y_i = +1 \quad (9)$$

$$f(x_i) = \langle w, x_i \rangle + b_i \geq -1 + \xi_i, y_i = -1 \quad (10)$$

Sınırın maksimum hale getirilmesine yanlış sınıflandırma hatalarının minimum hale getirilmesi arasındaki denge C düzenleme parametresi ($0 < C < \infty$) ile kontrol edilebilir[4]. Düzenleme parametresi C kullanıcı tarafından belirlenir. Parametrenin çok küçük veya çok büyük seçilmesi optimum hiper düzlemi doğru belirlemeyez.

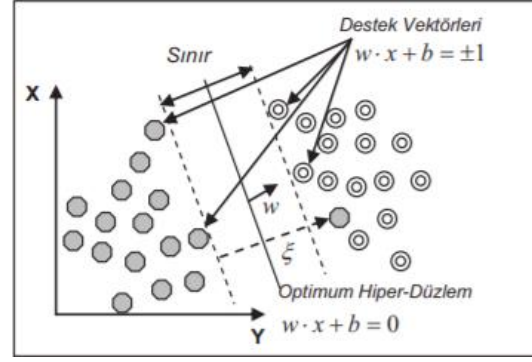
Yanlış sınıflandırma olasılığını düşürmek için optimizasyon problemi

$$\min \left[\frac{\|w\|^2}{2} + C \cdot \sum_{i=1}^r \xi_i \right] \quad (11)$$

$$y_i(w \cdot x_i + b) - 1 \geq 1 - \xi_i \quad (12)$$

$$\xi_i \geq 0 \text{ ve } i = 1, 2, \dots, N$$

şekline dönüşür.



Şekil 6. Doğrusal Olmayan DVM Optimum Hiper Düzlem

Gerçek yaşam problemlerinde veri setinin doğrusal bir hiper düzlemle ayrılması mümkün değildir. Bu nedenle çekirdek fonksiyonları kullanılır. Bu yöntem makine öğrenmesini arttırmaktadır[Taşkın]. Çekirdek fonksiyonu kullanarak doğrusal olarak ayrılmayan iki sınıflı problemin çözümü ile ilgili karar kuralı

$$f(x) = \text{sign}\left(\sum_{i=1}^k \alpha_i y_i \varphi(x) \cdot \varphi(x_i) + b\right) \quad (13)$$

şeklinde yazılabilir.[3]

Amaç sınıflandırma problemine göre kullanılacak çekirdek fonksiyonu ve bu fonksiyona ait parametrelerin optimum belirlenmesidir.

En çok kullanılan çekirdek yöntemleri

- Polinom Çekirdek
- Radyal Tabanlı Fonksiyon Çekirdek

Polinom Çekirdek Fonksiyonunda polinomun derecesindeki artış algoritmanın karmaşıklığını artırır. Bu da işlem süresini arttırmakta ve sınıflandırma doğruluğunu düşürmektedir.

Radyal Tabanlı Çekirdek Fonksiyonunda çekirdek boyutu (γ) olarak ifade edilen parametresindeki değişim sınıflandırma performansına etkisi en azdır. Gama değeri ne kadar yüksek olursa, eğitim veri setine o kadar uymaya çalışır.[1]

Çekirdek Fonksiyonu	Matematiksel İfade	Parametre
Polinom	$K(x, y) = ((x \cdot y) + 1)^d$	Polinom derecesi(d)
Radyal Tabanlı Fonksiyon	$K(x, y) = e^{-\gamma \ x - x_i\ ^2}$	Çekirdek boyutu(γ)

Tablo 1.Çekirdek Fonksiyon

b.Lojistik Regresyon

Lojistik Regresyon bir ya da daha fazla bağımsız veri kümesini analiz etmek için kullanılan istatistiksel bir yöntemdir.[https] En az değişken kullanarak en iyi uyuma sahip olacak şekilde bağımlı değişkenler arasındaki ilişkiyi tanımlayan bir model kurmaktır. Lojistik Regresyon sınıflandırma analizlerinde çokça kullanılır.

Lojistik regresyonda sonuç değişkeni ikili veya çokludur. Lojistik Regresyon üçe ayrılır.Bu modellerden “İkili Lojistik Regresyon Modeli” kategorik bağımlı değişken ikili(Örn: cinsiyet:kadın-erkek) olduğu durumda kullanılır.”Multinomial Lojistik Regresyon Modeli” kategorik bağımlı değişkenin çok kategorili(Örn: Medeni Durum;evli-bekar-boşanmış) olduğu durumlarda kullanılırken; sıralı bir yapı var ise (Örn: Ölçek; az-orta-çok) “Sıralı(Ordinal)Lojistik Regresyon kullanılır.[14]

İkili Lojistik regresyon ikili kategorik bağımlı değişken sahiptir.Yani bağımlı değişkenimiz 1 veya 0 değerlerinden birini alır. Lojistik regresyonun olasılıklar üzerinden çalışır.[16]

Lojistik değişkenin kullanılmasının sebebi dağılımı normalleştirmektir.

Lojistik Regresyon Fonksiyonu,

$$p(x) = \frac{\exp(\beta_0 + \beta^T x)}{1 + \exp(\beta_0 + \beta^T x)} \quad (14)$$

$p(x)$ 'nin modelde yer alan $(\beta_0 + \beta_1)$ parametrelerine dönüştürülmesine Logit adı verilir. Burada p, karakteristik özelliğinin var olma olasılığıdır.

$$\text{logit}(p) = \ln\left(\frac{p(x)}{1-p(x)}\right) = (\beta_0 + \beta_1 x) \quad (15)$$

Logit(p) modeldeki parametrelerle doğrusaldır ve süreklidir. $p(x)$ artıkçalogit(p) değeri de artar. [15]

Bağımlı Değişken (y)	Bağımsız Değişken (x)	
	x = 1	x = 0
y = 1	$P(1) = \frac{e^{\beta_0 + \beta_1}}{1 + e^{\beta_0 + \beta_1}}$	$P(0) = \frac{e^{\beta_0}}{1 + e^{\beta_0}}$
y = 0	$1 - P(1) = \frac{1}{1 + e^{\beta_0 + \beta_1}}$	$1 - P(0) = \frac{1}{1 + e^{\beta_0}}$
TOPLAM	1	1

Şekil 6. Lojistik Regresyon Değişken Olasılık Tablosu

İkili sınıflı bağımlı değişkenin iki kategrosinin görülme olasılıklarının birbirine oranlanmasına “Odds” adı verilir.Odds oranı

$$w = \frac{p}{1-p} \quad (16)$$

$$x=1 \text{ için } w_1 = \frac{P(1)}{1-P(1)} \quad x=0 \text{ için } w_1 = \frac{P(0)}{1-P(0)} \quad (17)$$

Bu durumda Odds Oranı;

$$\Psi = \frac{w_1}{w_0} = \frac{\frac{P(1)}{1-P(1)}}{\frac{P(0)}{1-P(0)}} = \frac{\frac{e^{\beta_0 + \beta_1}}{1 + e^{\beta_0 + \beta_1}}}{\frac{e^{\beta_0}}{1 + e^{\beta_0}}} = \frac{e^{\beta_0 + \beta_1}}{e^{\beta_0}} = e^{\beta_1} \quad (18)$$

olacaktır.

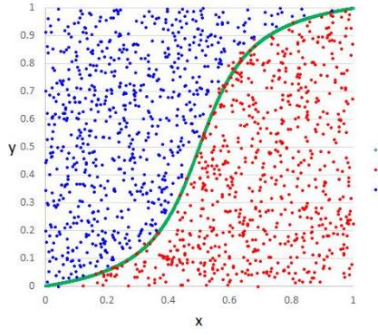
Bu iki sonuçlu bağımsız değişkenin lojistik regresyonu için odds oranıdır.Bu odds oranı için logit farkı ise

$$\ln(\psi) = \ln e^{\beta_1} = \beta_1 \quad (19)$$

katsayısına eşittir.[13]

İkili kategorili değişken arasındaki ilişkinin ölçüldüğü alanlar olan odds oranları hassastır ve lojistik regresyon analizinde önemli bir ölçüttür.[11]

Lojistik regresyonda tahmin, örnek değerlerin gözlem olasılığını en yükseğe çıkaran parametreleri seçer.



Şekil 8. Lojistik Regresyon

Lojistik Regresyon ile doğrusal regresyon arasındaki en belirgin fark lojistik regresyonda değişkenin ikili olmasıdır. Bunun yanında da bazı önemli farklarda bulunmaktadır.

Doğrusal regresyonda bağımlı değişken sürekli iken, lojistik regresyonda bağımlı değişken kesikli bir değer olmalıdır.

Doğrusal regresyonda bağımlı değişkenin değeri, lojistik regresyonda ise bağımlı değişkenin alabileceği değerlerden birinin gerçekleşme olasılığı kestirilir.

Doğrusal regresyonda bağımsız değişken çoklu normal dağılım gösterirken, lojistik regresyonun uygulanabilmesi için bağımsız değişkenlerin dağılımına ilişkin bir koşul yoktur.[12]

c. Rastgele Orman (Random Forest)

Rastgele orman algoritması ağaç tabanlı bir sınıflandırma algoritmasıdır.

Başlangıçta rastgele bir orman yaratılıyor. Yaratılan bu ağaçlar ile elde edilen sonuç arasında doğru orantı vardır[2].

Orman da yer alan her bir karar ağacı bootstrap tekniği ile veri setinden örneklem seçilmesi ve rastgele özellik seçilmesi ile oluşturulur. Bu yöntemde CART algoritması ile oluşturulur ve ağaçlar budanmaz. Bu algoritma “bilgi kazancını” kullanarak karar verir.[23]

Ağaç sayısındaki artış veri setinin türü ve büyüklüğüne ile bağlantılı olarak her zaman yüksek performans elde etmeyi sağlamaz.

Rastgele Orman algoritması ile Karar Ağacı algoritması arasındaki temel fark, Rastgele Ormanda kök düğümü bulma ve düğümleri bölme işlemlerinin rastgele çalışıyor olmasıdır.

Rastgele Orman Parametreleri

n_estimators :

Ormandaki ağaç sayısını temsil eder. Genellikle ağaç sayısı arttıkça öğrenme durumu artar. Bununla birlikte, çok sayıda ağaç eklenmesi eğitim sürecini önemli ölçüde yavaşlatabilir, bu nedenle optimum noktayı bulmak için bir parametre araştırması yapılmalıdır. Test performansı düşmeye başladığında ağaç sayısı arttırmayı durdurmak gereklidir[2].

max_feature :

Bir düğümü bölerken dikkate alınacak özelliklerin rasgele altkümelerinin boyutudur.

Bu boyut, auto veya sqrt olarak ayarlanırsa, özellik sayısının kareköküne ayarlanır.

Regresyon problemi için log2, sınıflandırma problemi için ise bu parametrenin 'sqrt' değeri alması akıllıcadır.

criterion :

Her bölmenin kalitesini ölçer. Gini veya Entropy olabilir. Entropy bilgi kazanımına dayanarak bölünmeyi sağlar.

$$E(S) = \sum_{i=1}^c -p_i \log_2 p_i \quad (18)$$

IV. ÖZET VE SONUÇLAR

	Train Accuracy	Accuracy	Precision	Recall
Rastgele Ağaç	100	77.08	81.15	79.43
Destek Vektör Makineleri	98.89	82.08	86.56	82.26
Lojistik Regresyon	94.64	70.83	75.53	74.46

Tablo 2. El Yazısından Cinsiyet Tahmini için Doğruluk değerleri

Yapılan testler sonucunda ortaya çıkan sonuçlar tatmin edicidir. Özellikle Destek vektör makinelerinin diğer makine öğrenme yöntemlerinden başarısız olarak ayrıldığı gözlemlenmektedir.

Destek Vektör makinelerinin güçlü yapısı bu veri setinde de kendini göstermeyi başarmıştır.

Lojistik Regresyon için aynı şeyleri söylemek yanlış olur. Bu derece karmaşık bir veri kümesine uygun bir model olmadığı açıkça görülmektedir.

KAYNAKLAR

- [1] Kavzoğlu, Taşkın, ve Çölkesen, İ., 2010, "Destek Vektör Makineleri ile Uydu Görüntülerinin Sınıflandırılmasında Kernel Fonksiyonlarının Etkilerinin İncelenmesi"
- [2] Y. Dong, B. Du, L. Zhang, "Target Detection Based on Random Forest Metric Learning", vol. 8, no. 4, pp. 1830-1838, 2015.
- [3] Osuna, E.E., Freund, R., Girosi, F., 1997, Support Vector Machines: Training and Applications, A.I. Memo No. 1602, C.B.C.L. Paper No. 144, Massachusetts Institute of Technology and Artificial Intelligence Laboratory, Massachusetts.
- [4] Cortes, C., Vapnik, V., 1995, Support-Vector Network, Machine Learning, 20(3): 273–297
- [5] Vapnik, V.N., 1995, The Nature of Statistical Learning Theory, Springer-Verlag, New York
- [6] Soman, K.P., Loganathan, R. and Ajay, V. (2011).Machine learning with SVM and other kernel methods. PHI Learning Pvt. Ltd., 486 s.
- [7] Gunn, S. R. (1998). Support vector machines for classification and regression. Technical Report, Faculty of Engineering, Science and Mathematics, School of Electronics and Computer Science.<http://users.ecs.soton.ac.uk/srg/publications/pdf/SVM.pdf>
- [8] Sevgi A,Şenol E."Destek Vektör Makineleriyle Sınıflandırma Problemlerinin Çözümü İçin Çekirdek Fonksiyonu Seçimi" Eskişehir Osmangazi Üniversitesi İİBF Dergi,Nisan 2014
- [9] <https://medium.com/@k.ulgen90/makine-%C3%B6%C4%9Frenimi-b%C3%B6%C3%BCm-4-destek-vekt%C3%B6r-makineleri-2f8010824054>
- [10]<https://veribilimcisi.com/2017/07/19/destek-vektor-makineleri-support-vector-machine/>
- [11]Allison, D. P., "Logistic Regression Using The SAS System 2nd ed.", SAS
- [12]Şibel Ç. Mahmut K.Akın Ç,Hüdaverdi B. "Lojistik Regresyon Analizinin İncilenmesi ve Diş Hekimliğinde Bir Uygulaması",Cumhuriyet Üniversitesi,2004
- [13]Zeynep Burcu Güner "Veri Madenciliğinde Cart ve Lojistik Regresyon Analizinin Yeri:İlaç Provizyon Sistemi Verileri Üzerinde Örnek Bir Uygulama,Sosyal Güvence Dergisi, 2014
- [14]Barak, A., Karahan, S. ve Saraçbaşı, O. (2005, Mayıs).Ordinal lojistik regresyon modelleri. 4. İstatistik Kongresi'nde sunulan bildiri, Belek-Antalya
- [15]Hüdaverdi Bircan "Lojistik Regresyon Analizi: Tıp Verileri Üzerine Bir Uygulama", Kocaeli Üniversitesi Sosyal Bilimler Enstitüsü Dergisi,2004
- [16]Pınar Bayrak,Logistic Regresyon Teori ve SPSS Çözümleri,2013
- [17]<https://veribilimcisi.com/2017/07/18/lojistik-regresyon/>
- [18]X. Liu, J. Tang, "Mass classification in mammograms using selected geometry and texture features and a new
- [19]SVM-based feature selection method", vol. 8, no. 3, pp. 910-920, 2014.
- [20] I. Guyon, A. Elisseeff, "An introduction to variable and feature selection", vol. 3, pp. 1157-1182, Mar 2003.
- [21]Y. Liu, F. Tang, Z. Zeng, "Feature selection based on dependency margin", vol. 45, no. 6, pp. 1209-1221, 2015.
- [22]A. E. Youssef, A. S. Ibrahim, A. L. Abbott, "Automated gender identification for arabic and english handwriting", (ICDP 2013), 2013.
- [23] Akman M,Genç Y,Ankaralı H ,“ Random Forest Yöntemi ve Sağlık Alanında Bir Uygulama” , Türkiye Klinikleri,2011