**Makale Künyesi**

Berna Arı1\*\*, Ali Arı2, İlknur Ucuz3, Filiz Çiledağ Özdemir4, Abdulkadir Şengür1 (2020), Grafik Tablet Kullanılarak Makine Öğrenmesi Yardımı ile El Yazısından Cinsiyet Tespiti, Fırat Üniversitesi Müh. Bil. Dergisi. 32(1), 243-252

**Anahtar kelimeler:** Cinsiyet sınıflandırma, el yazısı tanıma, makine öğrenmesi, kelime bölütleme, grafik tabletle el yazısı alımı

Bu makale, KA, NB, DVM, K-EYK sınıflandırıcılarıyla bireylere ait el yazısı örneklerinden cinsiyet tahmini yapmak amacıyla kullanan bir sistem oluşturmayı amaçlamaktadır. Yazının kişilere spesifik olması ve bireysel özellikler taşıması bilimin farklı alanlarında kullanımını gündeme getirmiştir. Bu duruma paralel, el yazısının karakterize edilmesini ve analizini gerçekleştiren grafoloji alanı gelişmiştir. Birçok disiplin, el yazısı özelliklerini kendi sahasında kullanımına yönelik çalışmalar gerçekleştirmişlerdir. Makalede de daha önce el yazısı analizi ile ilgili yapılan birçok araştırmadan bahsedilmiştir. Bu araştırmada de el yazısıyla cinsiyet tespiti, makine öğrenmesi disipliniyle bir araya getirilmiştir.

Araştırmada, el yazısı tanıma yapabilmek için ilk olarak genel hatlarıyla karakter tanıma çalışması yapılmıştır. Bu durumda karşımıza etkileşimsiz (çevrimdışı) ve etkileşimli (çevrimiçi) tanıma sistemleri çıkmaktadır. Çevrimdışı tanıma sistemleri, kâğıt üzerine yazılmış bir yazının sayısallaştırma adımlarını içerirken; çevrimiçi tanıma sistemleri, genellikle elektronik tabletler tarafından kalem hareketlerinin koordinatlarının elde edilmesiyle el yazısı ya da çizimlerin otomatik olarak algılanmasını sağlayan sistemlerdir

**Çalışmanın amacı ve veri seti**

Bu çalışmada önceden verilen belirli bir metnin 10 kişi tarafından tablete yazılması istendiği ve kişilerin özelliklerinin (cinsiyet, yaş, eğitim durumu, hangi elini kullandığı, ana dili vb.) not edildiği Marzinotto ve arkadaşları tarafından hazırlanan bir çevrimiçi el yazısı veri seti kullanılmıştır. Tasarlanan sistemde veri setindeki örneklerin boyutu arttırılmak istenmiştir. Bunun için el yazısı örnekleri, tüm metnin incelenmesi yerine veri setindeki 60 satırlık yazı kelime kelime bölütlenmiş ve 410 adet örnek elde edilmiştir.

**Öznitelik çıkarımı**

“Kalem ekran yüzeyine değmeden de beyaz noktalarla belirli aktif alanın biraz üzerindeki hareketleri algılayabilmektedir. Tablete kalemle Yazarken oluşturulan basınç hassasiyeti ilgili kalem ile ölçülebilmektedir. Tabletin basıncının yanında düzlem üzerindeki koordinatları ve yaptığı açı da yine alınabilen önemli değerler arasındadır” Bu duruma bağlı olarak analiz üç aşamada gerçekleşmiştir. Kalem havadayken, kalem tablet üzerindeyken ve yazı yazma halindeyken yani her iki durumun da kullanılmasıyla. Açıklanan kalemin havada olması durumunda; kalem hızı, ivmesi, yazarken oluşan sarsıntı hareketleri, eğim açısı, yazıdaki kavislenmeler, kalemin havada kalma oranı, kalemin yaptığı basınç değeri ve kalemin yükseklik açısı öznitelikleri elde edilmiştir. “Kalemin tablet üzerinde olması” ve “her iki durumda” ise kalemin hızı, ivmesi, sarsıntısı, eğim açısı, kavislenme öznitelikleri elde edilmiştir.

**Sınıflandırma**

Araştırmada dört farklı sınıflandırma yöntemi kullanılmıştır. Bunlar sırasıyla:

1. **Karar ağaçları:** Karar ağaçlarının amacı, girdi verilerinden çıkarılan basit karar kurallarını öğrenmek ve sonrasında, bir test örneğinin sınıf etiketini tahmin eden bir model oluşturmaktır.
2. **Naive Bayes:** Örneklerin her bir niteliğinin bağımsız olduğu ve her bir nitelik değerinin, diğer nitelikler hakkında bilgi içermediği temeline dayanmaktadır.
3. **Destek vektör makineleri:** İkili sınıflandırma probleminde doğrusal olarak ayrılabilen bir verisetinin olduğu düşünülürse, bu veri setini ayırabilen sonsuz sayıda hiper-düzlem vardır. DVM karar düzeyini oluştururken iki sınıfa olan uzaklığı maksimum yapmaya çalışır.
4. **K-en yakın komşu**: K-EYK yaklaşımı başlangıçta, bir mesafe fonksiyonuna sahip test örneği ve eğitim örnekleri arasındaki benzerlik ölçüsünü hesaplar. Daha sonra, hesaplanan benzerlik ölçüsüne dayanarak, test örneğinin en yakın komşuları belirlenir**.**

**Deneysel Sonuçlar ve Tartışmalar**

Bu çalışma; ilk olarak veri setinin alınmasını ve veri seti içiresindeki metinlerin kelime kelime bölütlenmesini içermektedir. Sonrasında yukarıda belirttiğimiz üç duruma(. Kalem havadayken, kalem tablet üzerindeyken, her iki durumda) göre öznitelik çıkarımı yapılmıştır

Yapılan cinsiyet sınıflandırma çalışmasında, kadın ve erkek el yazılarına ait 410 kelimeden oluşan örnek işleme alınmıştır. Sınıflandırmayı yapmak için veri setinin analiz işlemi 5 katlı çapraz doğrulama ile gerçekleştirilmiştir.

Yapılan araştırma sonucuna göre en yüksek cinsiyet sınıflandırma doğruluk oranı Kuadratik DVM ile %85,1 bulunmuştur. Ağırlıklandırılmış k-EYK ile %82,4, KA ile ortalama %73 doğruluk elde edilirken Çekirdek NB ile bu başarım oranı %79 bulunmuştur.

**Performans değerlendirme kriterleri:**

Makine öğrenmesi sınıflandırma modellerinin değerlendirilmesi maksadıyla hata matrisi yorumlanırken Gerçek Pozitif (GP), Gerçek Negatif (GN), Yanlış Pozitif (YP) ve Yanlış Negatif (YN) metrikleri işleme alınmıştır. Tasarlanan sistemde Hata Matrisine ait metriklerden Doğruluk, Duyarlılık ve Özgüllük değerleri hesaplanarak performans ölçülmek istenmiştir.

***Doğruluk (Accuracy):*** Doğru sınıflandırma değerinin toplam değere bölünmesi ile elde edilir

***Duyarlılık (Sensivity):*** Sistemin doğruları bilme konusundaki etkinliğidir. GP ‘nin YN ve GP değerleri toplamına bölümüyle bulunur ve bu değer, gerçekte bulmak istenilen cinsiyetin ne kadarının saptanabildiğidir

***Özgüllük (Specifity):*** Sistemde GN olan durumlara GN ve YP değerleri toplamının oranlanmasıyla bulunur ve gerçek durumu diğer cinsiyet olan bireylerin ne derece doğru tespit edildiğidir.

***Sonuçlar***

Bu çalışma, KA, NB, DVM, K-EYK sınıflandırıcılarıyla bireylere ait el yazısı örneklerinden cinsiyet tahmini yapmak amacıyla kullanan bir sistem önermektedir. Kullanılan bu sınıflandırıcılar kalemin havada olduğu, tablet üzerinde olduğu ve her iki durumdaki özniteliklerini (kalemin hızı, kalemin ivmesi, sarsıntı, eğim açısı, kavislenme, kalemin havada kalma oranı, kalemin basıncı yükseklik açısı) girdi olarak alır, sonrasında ise çıktı olarak cinsiyeti verir.

Sistem üzerinde, verilen el yazısı örneklerini cinsiyete göre tespit işlevinde %85,1 sınıflandırma elde edilmiştir. Önerilen yöntem, mevcut teknikler ve benzer sınıflandırıcı türleri ile karşılaştırıldığında yüksek bir başarım oranı göstermektedir. El yazısı özelliklerinin değerlendirildiği bu sistem ile farklı bireysel özellikler saptanabilecektir. Ayrıca multidisipliner gözlem ve izlem çalışmalarına katkı sunmuştur.