

# Denge Ölçeği Verisi Üzerinde Sınıflandırma Yöntemlerinin Karşılaştırılması

## Comparison of Classification Methods on Balance Scale

Ahmet ERKOÇ  
Elektrik-Elektronik Mühendisliği  
KTO Karatay Üniversitesi  
Konya, Türkiye  
[ahmeterkoc@gmail.com](mailto:ahmeterkoc@gmail.com)

**Özetçe**—Bu çalışmada UCI sitesinde yer alan verilerden Tim Hume'nin oluşturduğu Denge Ölçeği (Terazi Kefeleri) veri kümesi üzerinde sınıflandırma yöntemleri karşılaştırılması yapılmıştır. Destek Vektör Makineleri, Naive Bayes, K-en yakın komşu algoritması ve Multilayer Perceptron yöntemleri kullanılarak sınıflandırma yapılmış ve bulunan sonuçlar istatistiksel olarak karşılaştırılmıştır.

**Anahtar Kelimeler** — Denge Ölçeği, Naive Bayes, Destek Vektör Makineleri, K-en yakın komşu, Multilayer Perceptron

**Abstract**—In this study, we compared the classification methods on the dataset of the Balance Scale, which Tim Hume created from the data in the UCI site. Support Vector Machines, Naive Bayes, K-nearest neighbors algorithm and Multilayer Perceptron methods were used and the results were statistically compared.

**Keywords**—Balance Scale, Naive Bayes, Support Vector Machines, K-nearest neighbors, Multilayer Perceptron

### I. GİRİŞ (INTRODUCTION)

Sınıflandırma kavramı veri kümesinde bulunan çeşitli sınıflardaki verileri basitçe tabirle sınıflarına dağıtmaktır. Çeşitli algoritmalar, bu dağılımları verilen eğitim kümelerinden dağılım şeklini öğrenmeye çalışır ve dağılımları bu test verileri doğrultusunda doğru şekilde sınıflandırmaya çalışır. Bu sınıfların değerlerini de ( label ) olarak adlandırırız. Bu sayede eğitim ve test sırasında verinin sınıfını belirleriz.


Denge Ölçeği verileri Denge[B], Sol Kefe[L] ve Sağ Kefe[R] olarak sınıflandırılmıştır. Tablo 1'de Denge Ölçeği niteliklerinin değer aralıkları verilmiştir.

TABLO I. DENGE ÖLÇEĞİ ATTRIBUTE VE RANGE VALUES DEĞERLERİ  
(BALANCE SCALE DATASET ATTRIBUTE AND RANGE VALUES)

Sıra	Balance Scale dataset attribute and range values		
	Nitelik	Aralık Değerleri Min.	Aralık Değerleri Max.
1	Sol Ağırlık - Left Weight	1	5
2	Sol Mesafe - Left Distance	1	5
3	Sağ Ağırlık - Right Weight	1	5
4	Sağ Mesafe - Right Distance	1	5

### II. KULLANILAN YÖNTEMLER

#### A. Naive Bayes

Naive Bayes yöntemi, temelde Bayes Teoremine dayanan denetimli öğrenme  algoritma yapısına sahip olan veri kümeleri arasında bağımsızlık varsayan bir yöntemdir. Mevcut sınıflanmış örnekleri referans olarak, yeni örneği mevcut sınıflardan herhangiisine ait olma olasılığını hesaplar. Her

örnek aynı önem katsayısına sahiptir. Basite indirgenmiş varsayımlara rağmen bu yöntem mail spam filtreleme ve metin sınıflandırma gibi günümüzdeki şartlarda başarılı şekilde çalışır. Naive Bayes sınıflandırma teoremi (1) şu şekilde verilmektedir [1].

$$P(S_i) \prod_{K=1}^L P(X_K S_i) > P(S_j) \prod_{K=1}^L P(X_K S_j)$$

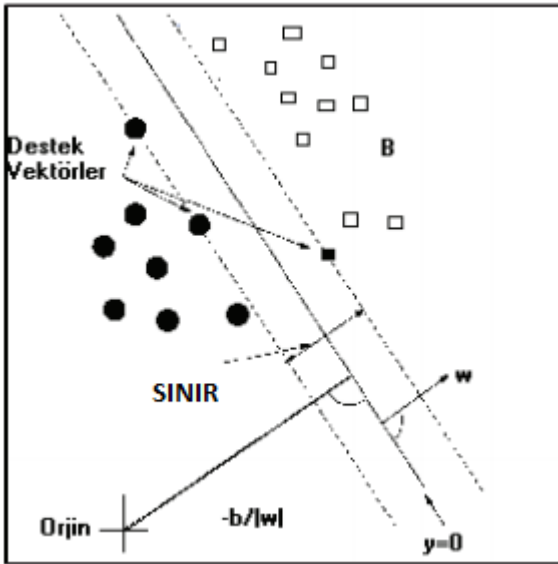
(1)  $P(S_i)$  ve  $P(S_j)$   $i$  ve  $j$  sınıflarının öncel olasılıklarıdır. Eldeki değerler kümesinden kolayca hesaplanır. Naive Bayes sınıflandırıcısının yeterli veriyle  $x$ 'in istatistik olarak bağımsız olması durumunda, yüksek başarılar elde edilir.

#### B. Destek Vektör Makineleri

Sınıflandırma alanında sık kullanılan bu basit yöntem oldukça da etkilidir. Doğrusal olarak ayırt edilebilen ya da edilemeyen veri kümeleri (DVM) ile sınıflandırılabilir. Bir düzlemde bulunan sınıflandırılacak iki grup arasında bir sınır çizilerek iki gruba ayırmak mümkündür. Bu sınır çizgilerini çizerken iki grup üyelerinin en uzağında kalan yerin seçimi önemlidir. İşte bu yöntem (DVM) sınırların düzlem üzerinde çizileceği doğru yeri belirler. Buradaki temel fikir, düzlemi ve boyutları birer özellik olarak düşünebilmektir. Her girişin bir özelliği olmuş, iki boyutlu düzlem üzerinde her girdinin farklı bir noktası oluşturulmuştur. Sınıfların arasındaki boşluğa ise tolerans demek mümkündür.

Bulduğu düzlemin denklemi (2) ile ifade edilir.

$$f(x) = (w, x) + b \quad (2)$$



Şekil 1. DVM Sınıflandırıcı Yapısı.

#### C. K-En Yakın Komşu Algoritması

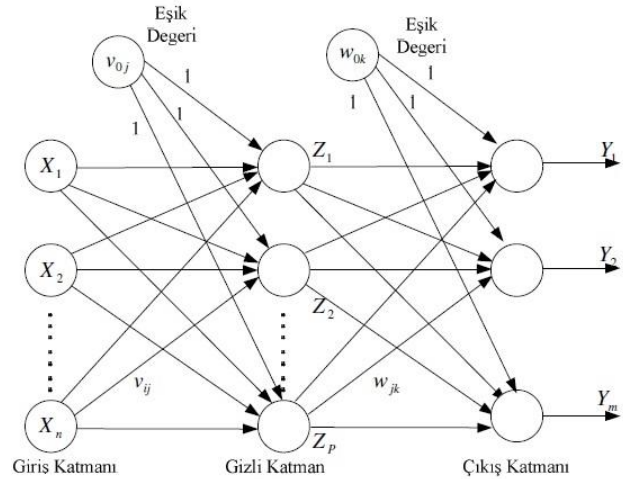
Araştırmacıların genellikle sinyaller ve örüntüleri sınıflandırmada tercih ettikleri K-En Yakın Komşu

algoritması, sezgisel ve basit bir denetimli öğrenme metodudur. Bu algoritma, test verisi ile eğitim verisini karşılaştırır. Test verisi bir örnektir, eğitim verisi ise temel verimizdir. Sınıflandırmak istediğimiz test verisine en yakın eğitim verisi (K-Komşu) tespit edilir. Sonrasında  $k$  tane komşu sınıflandırmadan en fazla hangisine benzediği belirlenir. Daha sonar sınıfın etiketi, sınıflandırılacak örneğe verilir [2].

$$d(x, x') = \sqrt{(x_1 - x'_1)^2 + (x_2 - x'_2)^2 + \dots + (x_n - x'_n)^2}$$

#### D. Multilayer Perceptron

Çok katmanlı ileri beslemeli yapay sinir ağı olarak da tanımlanan bu ağa kısaca MLP diyebiliriz. MLP ağlarında nöronlar katmanlar şeklinde oluşturulmuştur. İlk katmanında girdileri mevcuttur. Bu katmanda, çözümü arzu edilen bilgilerin YSA' ya girişini sağlamış oluruz. Son katman ise ağ içerisinde işlenmiş olan bilginin çıktı katmanıdır. Girdi ve çıktıların ortasındaki katmana gizli katman denilir. Gizli katman için herhangi şart sınırlama yoktur. Tek ya da birden çok olabilir. Bu durum MLP'nin durumuna göre değişiklik gösterir.



Şekil 2. Çok Katmanlı İleri Beslemeli Yapay Sinir Ağı

### III. DENEYSEL SONUÇLAR

Denge Ölçeği veri kümesinde sınıflandırma yaparken sınıflandırma yöntemleri kullanıldığını makalemizde bahsettik. Deneysel yöntemler kısmında yöntemlerimizi kullanırken elde ettiğimiz test seçeneklerinden bahsedeceğiz.

Weka programımızda Classifier; Naive Bayes, Classifier; Lazy IBK ve Classifier; MultiLayer Perceptron yöntemi kullanırken Test Option kısmında Cross-Validation aktif

olmuştur. Kısaca bahsedecek olursak; Cross-Validation, Folds metin içeriğindeki sayıyı kabul eder ve kullanır. 10 farklı sete ayırma işleminin ardından 9 tanesi ile eğitim 1 tanesi ile test yapar ve işlemini tamamlar.

Bir diğer test seçeneğimizden bahsecek olursak; Classifier; Function SMO kullanılırken aktif olan Percentage Split belirlenilen yüzde miktarına göre sınıflandırıcının ne kadar başarılı olduğunu test eder ve girilen yüzde miktarına göre bu işlemi yapar.

#### IV. SONUÇ

Bu çalışmada Denge Ölçeği Verisi Üzerinde Sınıflandırma Yöntemlerinin Karşılaştırılması, Naive Bayes, Destek Vektör Makineleri, K-En Yakın Komşu Algoritması ve Multilayer Perceptron yöntemleri kullanılarak gerçekleştirilmiştir. Bu yöntemler günümüzde birçok sınıflandırma alanında kullanılmakta ve başarılı sonuçlar elde edildiği görülmektedir.

YSA tercihinin kolay, hızlı ve tutarlı sonuçlar verdiği görülmektedir. Araştırmalarda bu yöntemler sıklıkla tercih edilmektedir.

Classifier:	Naive Bayes
Select Attributes:	CfsSybset Evaluator
Test Option:	Cross-Validation-10
Başarı Oranı:	%90.4

Classifier:	Function SMO
Select Attributes:	Attribute Evaluator
Test Option:	Percentage Split-(%66)
Başarı Oranı:	%88.32

Classifier:	Lazy IBK
Select Attributes:	Attribute Evaluator
Test Option:	Cross-Validation-10
Başarı Oranı:	%85.5

Classifier:	MultiLayer Perceptron
Select Attributes:	Attribute Evaluator
Test Option:	Cross-Validation-10
Başarı Oranı:	<b>%92.1</b>

Aldığımız sonuçlara bakıldığında sınıflandırma yöntemlerinden en başarılısı %92.1 ile Multilayer Perceptron olarak görülmektedir. Diğer yöntem ve başarı miktarları tabloda verilmiştir. Ayrıca bir sonraki çalışmada veri madenciliği tekniklerinden C5.0 algoritması, doğrusal ve doğrusal olmayan regresyon modelleri yeniden karşılaştırılma düşünülmektedir.

#### KAYNAKLAR

- [1] Elçi, G. Altunsoy A., İkizler N. EEG "İşaretlerinde Sınıflandırma Yöntemlerinin İstatistiksel Olarak Karşılaştırılması" , 25. IEEE Sinyal İşleme ve İletişim Uygulamaları Kurultayı
- [2] Murugappan, M., Ramachandran, N., Sazali, Y. (2010). Classification of Human Emotion from EEG Using Discrete Wavelet Transform. Journal of Biomedical Science and Engineering, 3(04), 390.