



## Yüzüncü Yıl Üniversitesi Fen Bilimleri Enstitüsü Dergisi

<https://dergipark.org.tr/tr/pub/yyufbed>



Araştırma Makalesi

### Ar-Ge Teşviklerine Yönelik Sınıflandırma Problemleri için Çok Kriterli Karar Verme Yöntemleriyle Uzlaşık Karar Verme Nasıl Sağlanır?

Gözde ULU METİN<sup>1\*</sup> Özlem TÜRKŞEN<sup>2</sup>

<sup>1</sup>ASELSAN A.Ş., 06200, Ankara, Türkiye

<sup>2</sup>Ankara Üniversitesi, Fen Fakültesi, İstatistik Bölümü, 06560, Ankara, Türkiye

\* Sorumlu yazar e-posta: [gozdeulumetin@gmail.com](mailto:gozdeulumetin@gmail.com)

**Öz:** Çok Kriterli Karar Verme (ÇKKV) yöntemleri, veri keşfi sürecinin ilk adımından son adımına kadar her aşamada karar destek sistemlerinde aktif rol oynayan bilimsel yaklaşımları içerir. Bu çalışmada, ÇKKV yöntemleri, sınıflandırma problemlerinde makine öğrenmesi algoritmaları ile elde edilen sonuçların değerlendirilmesinde uzlaşık çözüm elde edilmesi amacıyla uygulanmıştır. Veriye ait her bir özellik farklı önem düzeylerine sahip olduğundan bu özelliklerin ağırlıklandırılmasında da ÇKKV yöntemleri kullanılmaktadır. Buna göre, çalışmada, özellik ağırlıklandırması amacıyla AHP, Entropi ve CRITIC yöntemleri kullanılmıştır. Ayrıca, ağırlıklandırılmış verinin sınıflandırılması amacıyla Lojistik Regresyon,  $k$ -En Yakın Komşu algoritması, Destek Vektör Makineleri ve Rasgele Orman makine öğrenmesi sınıflandırma algoritmaları uygulanmıştır. Sınıflandırma algoritmalarının performansının değerlendirilmesi için Doğruluk, Kesinlik, Duyarlılık, F1-Skor ve AUC gibi çeşitli performans ölçütleri kullanılmıştır. Bu ölçütler kriter olarak kabul edilerek bir karar matrisi oluşturulmuştur. Her bir kriter, IDOCRIW yöntemi ile ağırlıklandırılmıştır. Uygulanan sınıflandırma yöntemlerinin sıralamaları, TOPSIS, EDAS, CODAS ve MABAC gibi çeşitli ÇKKV yöntemleri ile hesaplanmıştır. Uygulama kapsamında Ar-Ge teşvikleri alanından bir sınıflandırma problemi ele alınmıştır. Tüm hesaplamalar Python 3.11.3. PyDecision kütüphanesi kullanılarak gerçekleştirilmiş ve elde edilen sonuçlar ayrıntılı biçimde analiz edilip tartışılmıştır. AHP, Entropi ve CRITIC yöntemleri ile ağırlıklandırılmış sınıflandırma algoritmalarının performansları farklı ÇKKV yöntemleri ile sıralandığında elde edilen tüm sonuçlarda Rasgele Orman algoritmasının öncelikli tercih edildiği,  $k$ -En Yakın Komşu algoritmasının son sırada tercih edilebileceği bilgisine ulaşılmıştır.

**Anahtar Kelimeler:** Ar-Ge Teşvikleri, Çok Kriterli Karar Verme (ÇKKV) Yöntemleri, Makine Öğrenmesi Sınıflandırma Algoritmaları, Performans Ölçütleri, Sınıflandırma Problemi

### How to Achieve Consensus for Classification Problems about R&D Incentives with Multi-Criteria Decision Making Methods?

**Abstract:** Multi-Criteria Decision Making (MCDM) methods include scientific approaches that play an active role in decision support systems from the first step to the last step of the data discovery process. In this study, the MCDM methods are used to reach a consensus on the results obtained using machine learning algorithms for classification problems. As each data feature has a different level of importance, MCDM methods are also used to weight these features. In this study, AHP, Entropy and CRITIC methods are used for feature weighting. Additionally, the following machine learning classification algorithms were applied to classify the weighted data: Logistic Regression,  $k$ -Nearest Neighbours, Support Vector Machines and Random Forest. Various performance measures, such as Accuracy, Precision, Sensitivity, F1-Score and AUC, were used to evaluate the performance of the classification methods. These criteria were adopted and a decision matrix was created. Each criterion was weighted using the IDOCRIW method. The rankings of the classification methods applied were calculated using

Gönderilme Tarihi: 02.07.2025

Kabul Tarihi: 22.09.2025

**Nasıl atıf yapılır:** Ulu Metin, G., & Türkşen, Ö. (2025). Ar-Ge teşviklerine yönelik sınıflandırma problemleri için çok kriterli karar verme yöntemleriyle uzlaşık karar verme nasıl sağlanır?. *Yüzüncü Yıl Üniversitesi Fen Bilimleri Enstitüsü Dergisi*, 30(3), 878-896. <https://doi.org/10.53433/yyufbed.1731123>

various MCDM methods, such as TOPSIS, EDAS, CODAS and MABAC. The application considers a classification problem in the field of R&D incentives. All calculations were performed using the Python 3.11.3. PyDecision library. The results were analysed and discussed in detail. When the performances of the classification algorithms weighted with AHP, Entropy and CRITIC methods are ranked using different MCDM methods, it is concluded that the Random Forest algorithm is the preferred among all of the obtained results while the  $k$ -Nearest Neighbours algorithm is the least preferred one.

**Keywords:** R&D Incentives, Multi Criteria Decision Making Methods (MCDM), Machine Learning Classification Algorithms, Performance Metrics, Classification Problem

## 1. Giriş

Bilgi teknolojilerindeki gelişmeler ile birlikte dijital ortamlarda büyük miktarda veri üretilebilmekte ve saklanabilmektedir. Depolanan ham verilerden anlamlı ve kullanılabilir bilginin elde edilebilmesi için ham verinin işlenmesi süreci kritik öneme sahiptir (Ulu Metin, 2024). Bu nedenle, veri analizi süreçlerinin temelini oluşturan bilgi keşfi, oldukça önem kazanmıştır.

Fayyad ve ark. (1996) tarafından yapılan çalışmada, verilerin, bilgi keşfinden yararlanılarak etkin biçimde analiz edilmesiyle karar verme süreçlerine katkı sağlayacak değerlerde bilgiler üretildiği vurgulanmaktadır. Veriden bilgi keşfinin ön işleme aşamasında, veri temizleme, bütünleştirme, boyut indirgeme ve veri seçimi gibi işlemler istatistiksel yaklaşımlara dayalı olarak gerçekleştirilmektedir. Bu sayede eksik değerler tamamlanmakta, tutarsız ve anlamsız veriler çıkartılmaktadır. Gerekli durumlarda, uygun dönüşüm teknikleriyle veri dönüşümü yapılarak veri, analiz sürecine hazır hale getirilmektedir. Betimsel istatistikler ve veri görselleştirme yöntemleri ile veri setindeki değişkenlerin temel istatistiksel özellikleri ortaya konularak verinin genel yapısının anlaşılması sağlanmaktadır.

Veri setindeki her bir değişken veriye ait özellikleri ifade eder. Betimsel istatistikler hesaplanmadan önce veri setindeki değişken türleri incelenir.  $m$  sayıda bağımsız değişken, bir bağımlı değişkenden oluşan  $n$  gözlemlili veri seti örneği Çizelge 1’de görülmektedir.

Çizelge 1. Çok değişkenli bir veri seti örneği

| No. | Bağımsız değişkenler |          |     |          | Bağımlı değişken |
|-----|----------------------|----------|-----|----------|------------------|
|     | $X_1$                | $X_2$    | ... | $X_m$    | $Y$              |
| 1   | $x_{11}$             | $x_{12}$ | ... | $x_{1m}$ | $y_1$            |
| 2   | $x_{21}$             | $x_{22}$ | ... | $x_{2m}$ | $y_2$            |
| ⋮   | ⋮                    | ⋮        | ⋮   | ⋮        | ⋮                |
| $n$ | $x_{n1}$             | $x_{n2}$ | ... | $x_{nm}$ | $y_n$            |

Burada, bağımsız değişkenler  $X = [x_{ij}]_{n \times m}$  ve bağımlı değişken  $Y = [y_i]_{n \times 1}$  biçimindedir,  $i = 1, 2, \dots, n$ ,  $j = 1, 2, \dots, m$ . Veri setinin bir sınıflandırma problemine ilişkin olması durumunda, Çizelge 1’de görülen bağımlı değişken kategorik değerler alırken bağımsız değişkenler nicel ve/veya kategorik değerlerden oluşabilir. Veri setindeki her bir değişken farklı önem derecelerine sahip olabilir. ÇKKV yöntemlerinden yararlanılarak öznel ya da nesnel yaklaşımlarla veri setindeki değişkenlerin önem ağırlıkları  $w = [w_1 \ w_2 \ \dots \ w_m]$  hesaplanır. Elde edilen önem ağırlıkları kullanılarak yeni bağımsız değişkenler  $\zeta_i = w_i X_i$ ,  $i = 1, 2, \dots, m$  biçiminde hesaplandığında Çizelge 2’deki ağırlıklandırılmış veri seti elde edilir.

Çizelge 2. Çok değişkenli bir veri setinin ağırlıklandırılması

| No. | Bağımsız değişkenler |              |     |              | Bağımlı değişken |
|-----|----------------------|--------------|-----|--------------|------------------|
|     | $\zeta_1$            | $\zeta_2$    | ... | $\zeta_m$    | $Y$              |
| 1   | $\zeta_{11}$         | $\zeta_{12}$ | ... | $\zeta_{1m}$ | $y_1$            |
| 2   | $\zeta_{21}$         | $\zeta_{22}$ | ... | $\zeta_{2m}$ | $y_2$            |
| ... | ...                  | ...          | ... | ...          | ...              |
| $n$ | $\zeta_{n1}$         | $\zeta_{n2}$ | ... | $\zeta_{nm}$ | $y_n$            |

Ön işleme süreci tamamlanan ağırlıklandırılmış veri setine, makine öğrenmesi sınıflandırma algoritmaları uygulanarak öngörü modeli oluşturulur. Sınıflandırma algoritmalarının değerlendirilmesinde performans ölçütleri kullanılır. Hesaplanan performans değerlerine göre sınıflandırma algoritmalarının tercih sıralaması ÇKKV yöntemleri ile belirlenir. [Rafiei-Sardooi ve ark. \(2021\)](#) çalışmalarında, kentsel sel riski değerlendirmesinde ÇKKV ve makine öğrenmesi algoritmalarını birleştirerek hibrit bir yaklaşım sunmuşlardır. Çalışmada sel tehlikesini modellemek için Destek Vektör Makineleri, Rasgele Orman ve Desteklenmiş Regresyon Ağaçları kullanılmıştır. Model performansları AUC değerleri ile ölçülmüş olup Rasgele Orman diğer algoritmalarından daha iyi performans göstermiştir. [Xie & Zhang \(2023\)](#) çalışmalarında, birçok kategorik değişkene sahip gerçek bir veri setindeki değişken seçimi için TOPSIS kullanmıştır. Modelin sonucunu etkileyen en önemli değişkenler belirlenerek makine öğrenmesi algoritmalarından Karar Ağacı, Rasgele Orman ve XGBoost uygulanarak performansları hesaplanmıştır. Önerilen yaklaşımın XGBoost yönteminden daha iyi performansla Rasgele Orman algoritmasına ise benzer bir performans gösterdiği sonucuna ulaşılmıştır.

Bu çalışmada, makine öğrenmesi sınıflandırma algoritmalarına karar verilmesinde ağırlıklandırma ve sıralama amacıyla ÇKKV yöntemleri kullanılmıştır. Önerilen yaklaşımın Ar-Ge teşvikleri alanındaki gerçek veri seti üzerinde uygulanabilirliği gösterilmiştir. Bu çalışma, ÇKKV yöntemlerinin hem veri ön işleme sürecindeki ağırlıklandırma aşamasında ve hem de makine öğrenmesi algoritmalarının sıralama aşamasında kullanılmasıyla bütünlük ve uzlaşık bir karar verme yaklaşımı sunarak literatüre özgün bir katkı sağlamaktadır.

Çalışmanın ikinci bölümünde, değişkenleri ağırlıklandırılmış veri setine uygun olan makine öğrenmesi sınıflandırma algoritmalarından Lojistik Regresyon,  $k$ -En Yakın Komşu algoritması, Destek Vektör Makineleri ve Rasgele Orman'ın teorik açıklamaları, uygulama adımları ve ayarlanabilir parametreleri ile sınıflandırma performans ölçütleri hakkında bilgi verilmiştir. Çalışmanın üçüncü bölümünde ÇKKV yöntemleri ağırlıklandırma (AHP, Entropi ve CRITIC) ve sıralama (TOPSIS, CODAS, EDAS ve MABAC) başlıkları altında değerlendirilerek bu yöntemlere ilişkin uygulama adımları verilmiştir. Çalışmanın dördüncü bölümünde, önerilen yaklaşımın Ar-Ge teşvikleri alanındaki gerçek veri seti üzerinde uygulanabilirliği gösterilmiştir. Çalışmanın beşinci bölümünde ise gerçek veri setine ilişkin elde edilen sonuçlar yorumlanmıştır.

## 2. Makine Öğrenmesi Sınıflandırma Algoritmaları

Makine öğrenmesi sınıflandırma algoritmaları, verideki gizli örüntüleri ortaya çıkararak gözlemlerin hangi sınıfa ait olduğunu tahmin edilmesini sağlayan denetimli öğrenme yöntemleridir. Sınıflandırma algoritmalarının uygulanabilmesi için, her bir algoritmaya ilişkin ayarlanabilir parametrelerin uygun biçimde belirlenmiş olması gerekir. Ayarlanabilir parametrelerin seçimi, algoritmaların sınıflandırma performansında etkilidir. Bağımlı değişkenin kategorik ve bağımsız değişkenlerin sürekli ve/veya kategorik değişkenlerden oluştuğu veri setleri için aşağıdaki sınıflandırma algoritmaları kullanılmaktadır.

### 2.1. Lojistik regresyon

Lojistik regresyon (LR), veri setindeki gözlemlerin belirli bir sınıfa ait olup olma olasılığını hesaplayarak yeni gözlemleri sınıflandıran bir makine öğrenmesi algoritmasıdır.  $Y_i \in \{0,1\}$  iken  $\{(X_i, Y_i)\}_{i=1}^n$  etiketli bir veri seti olsun. Bu veri setinde, bağımlı değişken kategorik olduğundan bağımlı

değişkenin olasılığı hesaplanıp logit dönüşüm uygulanarak Çizelge 2’deki ağırlıklandırılmış veri setine ilişkin

$$f_{LR}(\zeta_i) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 \zeta_{i1} + \beta_2 \zeta_{i2} + \dots + \beta_m \zeta_{im})}}, \quad i = 1, 2, \dots, n \quad (1)$$

LR modeli oluşturulur. Eşitlik (1) bir olasılık değeri olup eşik değeri 0.5 alındığında

$$\hat{y}_i = \begin{cases} 1, & f(\zeta_i) \geq 0.5 \\ 0, & f(\zeta_i) < 0.5 \end{cases} \quad (2)$$

biçiminde sınıflandırılır. LR ile sınıflandırma problemine ilişkin optimizasyon modeli

$$\begin{aligned} \max_{\beta_0, \beta_1, \dots, \beta_m \in \mathbb{R}} \quad & \ln L(\beta_0, \beta_1, \dots, \beta_m) \\ g(\beta) = c \sum_{j=1}^m |\beta_j|^t \geq 0, \quad & t = 1, 2, \quad c \geq 0 \end{aligned} \quad (3)$$

biçiminde tanımlanır. Burada,  $\beta = [\beta_0 \quad \beta_1 \quad \dots \quad \beta_m]$  model parametre vektörüdür. (3) ifadesi ile tanımlı eşitsizlik kısıtlı doğrusal olmayan optimizasyon probleminin çözümü Kuhn-Tucker koşullarından yararlanılarak yapılır. LR’nin ayarlanabilir parametreleri, kısıtlama yapılacak yöntemi (Lasso ( $t = 1$ ), Ridge ( $t = 2$ ), None ( $c = 0$ )) belirleyen ceza değeri (penalty) ve kısıtlama oranı ( $c$ )’dır (Geron, 2017).

## 2.2. k-En yakın komşu algoritması

k-En Yakın Komşu (kNN) Algoritması, sınıfı belli olan gözlemler ile yeni gelecek gözlem arasındaki uzaklıklar hesaplanarak yeni gözlemin sınıflandırıldığı bir makine öğrenmesi algoritmasıdır. Hesaplanan en küçük uzaklığa sahip  $k$  sayıda gözlemin sınıfları içerisinde çoğunluk olan sınıf, yeni gözlem değerinin sınıfı olarak tahmin edilir (Sorhun, 2021). kNN algoritmasında uzaklıkların hesaplanmasında Öklid uzaklık formülü yaygın olarak kullanılır. Çizelge 2’de verilen ağırlıklandırılmış veri setine ilişkin gözlemler arasındaki uzaklık

$$f_{kNN}(\zeta_i, \zeta_j) = \sqrt{\sum_{t=1}^m (\zeta_{it} - \zeta_{jt})^2}, \quad i, j = 1, 2, \dots, n, \quad i \neq j \quad (4)$$

biçiminde yazılır. Burada, gözlemler arasındaki en küçük uzaklığın elde edilmesi amaçlandığından kNN sınıflandırma problemine ilişkin optimizasyon modeli

$$\min f_{kNN}(\zeta_i, \zeta_j) \quad (5)$$

olarak tanımlanır. (5) ifadesinde verilen doğrusal olmayan optimizasyon problem çözümü için türeve dayalı optimizasyon yöntemleri kullanılır (Türkşen, 2023). kNN algoritması ile sınıflandırmada  $k$  ayarlanabilir parametresinin belirlenmesi önemlidir.

## 2.3. Destek vektör makineleri

Destek Vektör Makineleri (SVM) veriyi sınıflamak için uygun bir fonksiyonu kullanan makine öğrenmesi algoritmasıdır.  $Y_i \in \{-1, 1\}$  iki sınıflı bağımlı değişken olmak üzere  $\{(X_i, Y_i)\}_{i=1}^n$ , etiketli bir veri seti olsun.  $Y_i = -1$  ve  $Y_i = 1$  olarak etiketlenmiş iki sınıf  $H$  hiperdüzlemi

$$H \text{ düzlemi} : \langle \beta, \zeta \rangle + \beta_0 = 0 \quad (6)$$

ile ayrılır.  $H$  düzlemi optimal hiper düzlem olup  $H$  düzlemine kenar payı uzaklıkları  $d$  olan  $H_1$  düzlemi ve  $H_2$  düzleminin denklemleri Eşitlik (6) kullanılarak

$$\begin{aligned} H_1 \text{ düzlemi: } \langle \beta, \zeta \rangle + \beta_0 - d &= 0 \\ H_2 \text{ düzlemi: } \langle \beta, \zeta \rangle + \beta_0 + d &= 0 \end{aligned} \quad (7)$$

biçiminde yazılır. Sınıflandırma sırasında, karşı sınıfta mümkün olan en az sayıda farklı sınıfa ait gözlemin bulunması ve  $H_1$  düzlemi ile  $H_2$  düzlemi üzerinde en az birer destek vektörünün yer alması beklenir. Gözlemler sınıflandırılırken kenar payı uzunluğunun en büyük olması amaçlanır (Metlek & Kayaalp, 2020). SVM ile

$$\begin{aligned} \min_{\beta \in \mathbb{R}^n} f(\beta) &= \frac{1}{2} \langle \beta, \beta \rangle \\ g(\beta, \beta_0) &= y_i (\langle \beta, \zeta_i \rangle + \beta_0) - 1 \geq 0, \quad i = 1, 2, \dots, n \end{aligned} \quad (8)$$

olarak tanımlanan problemin optimizasyonu için Kuhn-Tucker koşullarından yararlanılır ve problem çözülür. Çizelge 2’de verilen ağırlıklandırılmış veri setine ilişkin SVM ile yeni gelen gözlemler

$$f_{SVM}(\zeta) = \text{sign}(\beta^T \zeta + \beta_0) \quad (9)$$

biçiminde sınıflandırılır. SVM, doğrusal ayırımın sağlanamadığı ya da fazla sayıda değişken bulunduğu durumlarda, çekirdek fonksiyonlarını kullanır (Uğuz, 2021). Çekirdek fonksiyonu olarak

$$K(x_i, x_j) = \langle \phi_{x_i}, \phi_{x_j} \rangle = (\langle x_i, x_j \rangle)^2 \quad (10)$$

biçiminde tanımlı fonksiyon kullanılır. Yaygın olarak kullanılan çekirdek fonksiyonları doğrusal, polinom, dairesel tabanlı ve sigmoid fonksiyonlarıdır. SVM için ayarlanabilir parametreler, kısıtlama oranı ( $c$ ), çekirdek fonksiyonları ve  $c$  oranını düzenleyici değer olan gamma’dır.

#### 2.4. Rasgele orman algoritması

Rasgele Orman (RF), birden çok sayıda karar ağacının birleşmesiyle oluşan bir makine öğrenmesi algoritmasıdır. RF, her ağaçtan tahminleri toplayarak çoğunluk oylaması ile tahmin eder. Ağaç sayının artmasıyla modelin performansı artar. RF tüm değişkenleri kullanmak yerine her bir düğümde Gini ya da Entropi indeksi kullanarak değişken seçimi yapar.  $B$  oluşturulan ağaç sayısı olmak üzere  $b$ . ağacın sınıf tahmini  $\hat{C}_b(x)$  ‘dir,  $b = 1, 2, \dots, B$ . Çizelge 2’de verilen ağırlıklandırılmış veri setine ilişkin RF modeli

$$\hat{C}_{RF}^B(\zeta) = \text{çoğunluk oylaması} \left\{ \hat{C}_b(\zeta_i) \right\}_1^B, \quad i = 1, 2, \dots, n \quad (11)$$

biçiminde yazılır. Hata oranı minimum yapılarak ağaç performansının değerlendirilir (Arlı ve ark., 2022). Buna göre, RF ile sınıflandırma problemine ilişkin optimizasyon modeli

$$\min_{\zeta \in \mathbb{R}^n} E_{OOB} = \frac{1}{n} \sum_{B=1}^B I(y_i - \hat{C}_{RF}^B(\zeta)) \quad (12)$$

biçiminde tanımlı kısıtsız doğrusal olmayan optimizasyon problemi olarak tanımlanır ve çözülür. RF için ayarlanabilir parametreler; maksimum derinlik, maksimum değişken, minimum yaprak örneği, minimum örneklem, oluşan ağaç sayısı ( $b$ ) ve indeks hesaplama yöntemi (Gini, Entropi)’dir.

Sınıflandırma algoritmalarıyla elde edilen modellerin tahmin performansının değerlendirilebilmesi amacıyla çeşitli performans ölçütleri hesaplanır. Her bir sınıflandırma algoritmasına ait model performansını iyileştiren ayarlanabilir parametre değerlerinin belirlenmesine geçilir (Geron, 2021).

Sınıflandırma performans ölçütlerinin hesaplanabilmesi için öncelikli olarak veri seti, %80'i eğitim ve %20'si test veri setleri olmak üzere ikiye ayrılır. Eğitim veri seti kullanılarak sınıflandırma algoritması eğitilir ve model oluşturulur. Elde edilen modelin performansını artırmak amacıyla, uygulanan yönteme ait ayarlanabilir parametrelerin optimizasyonu gerçekleştirilir. Ancak, modelin eğitilmiş olması her zaman verinin tam olarak öğrenildiği anlamına gelmemekte, öğrenilen bilginin test verisine genellenememesi durumu ortaya çıkabilmektedir. Bu durum, aşırı öğrenme (overfitting) problemi olarak tanımlanmaktadır. Bu tür problemleri önlemek amacıyla modelin eğitim verisine karşı duyarlılığı çapraz doğrulama yöntemiyle sağlanır. Çapraz doğrulamada, veri seti eşit  $k$  sayıda alt gruba (katlara) ayrılır. Her seferinde bu gruplardan biri doğrulama, kalanlar ise eğitim amacıyla kullanılır. Bu işlem her grup için tekrarlanır ve her seferinde doğrulama hatası hesaplanır. Tüm doğrulama hatalarının ortalaması alınarak  $k$ -kat çapraz doğrulama sonucu elde edilir. Modelin başarımını artırmak için her algoritmaya ait farklı ayarlanabilir parametrelerle çeşitli model denemeleri yapılır. Bu denemeler ızgara arama yöntemiyle sistemli bir şekilde gerçekleştirilir ve en iyi sonucu veren parametre kombinasyonu seçilir. Modelin genel performansı, daha önce hiç kullanılmamış olan test veri setiyle değerlendirilir. Performans ölçütlerinin hesaplanmasında Çizelge 3'te verilen karışıklık matrisi kullanılır.

Çizelge 3. İkili sınıflandırma için karışıklık matrisi

|                        | <b>Tahmin: Pozitif</b> | <b>Tahmin: Negatif</b> |
|------------------------|------------------------|------------------------|
| <b>Gerçek: Pozitif</b> | Doğru Pozitif (DP)     | Yanlış Negatif (YN)    |
| <b>Gerçek Negatif</b>  | Yanlış Pozitif (YP)    | Doğru Negatif (DN)     |

Çizelge 3'te verilen karışıklık matrisi kullanılarak Doğruluk, Duyarlılık, Kesinlik,  $F_1$ -Skor ve AUC ölçütü hesaplanır. Bu matris, modelin tahmin ettiği sınıflarla gerçek sınıflar arasındaki ilişkiyi, DP, YP, DN ve YN olmak üzere dört kategoriye ayrılan gözlem sayılarıyla ifade eder. Karışıklık matrisinde, her sınıfa ait gözlem sayılarının birbirine yakın olması, veri setinin dengeli bir dağılıma sahip olduğunu gösterir. Doğruluk, doğru olarak sınıflandırılmış örneklerin toplam örnek sayısına oranı olarak

$$\text{Doğruluk} = \frac{DP + DN}{DP + DN + YP + YN} \quad (13)$$

biçiminde tanımlanır. Bu ölçüt, dengeli dağılmayan veri setlerinde yeterli bir performans ölçütü değildir (Hossin & Sulaiman, 2015). Duyarlılık, doğru olarak sınıflandırılmış pozitif örnek sayısının gerçekteki pozitif örnek sayısına oranı olup

$$\text{Duyarlılık} = \frac{DP}{DP + YN} \quad (14)$$

biçimindedir. Kesinlik, pozitif olarak tahmin edilen değerlerin ne oranda doğru olarak tahmin edildiğini belirtir ve

$$\text{Kesinlik} = \frac{DP}{DP + YP} \quad (15)$$

hesaplanır.  $F_1$ -Skor, kesinlik ve duyarlılık ölçütünün harmonik ortalaması alınarak

$$F_1 - \text{Skor} = 2 \frac{(\text{Duyarlılık})(\text{Kesinlik})}{(\text{Duyarlılık}) + (\text{Kesinlik})} \quad (16)$$

biçiminde hesaplanır. İkili sınıflandırma problemlerinde ve dengeli dağılıma sahip olmayan veri setlerinin değerlendirilmesinde doğruluktan daha iyi bir performans göstermektedir (Joshi, 2002; Uğuz, 2021). AUC ölçütü ise, ROC eğrisinin altında kalan alan toplamı ifade eden bir performans ölçütüdür (Kumar ve ark., 2020).

### 3. Çok Kriterli Karar Verme Yöntemleri

ÇKKV yöntemleri, veriye dayalı objektif karar vermeyi sağlayan yöntemlerdir. Ağırlıklandırma ve sıralama başlıklarında değerlendirilebilir.

#### 3.1. Ağırlıklandırma

Farklı önem derecesine ve ağırlığa sahip değişkenlerin olduğu bir veri setindeki önem ağırlıklarının belirlenmesinde ÇKKV yöntemlerinden yararlanılır. Ağırlıklandırmada kullanılan bu yöntemler öznel ya da nesnel olabilir. Öznel ağırlıklandırmada AHP, nesnel ağırlıklandırmada Entropi ve CRITIC yöntemleri kullanılabilir.

##### 3.1.1. AHP yöntemi

AHP (Analytic Hierarchy Process) yöntemi, sözel ifadeleri sayısal değerler ile ifade ederek karşılaştırma yapan ÇKKV yöntemlerinden biridir. AHP’de, Saaty (1977) tarafından önerilmiş olan karşılaştırma ölçeği ile öznel değerlendirmeler, matematiksel olarak ifade edilir. AHP uygulama adımları aşağıdaki gibidir.

**Adım1:** Çizelge 1’de verilen  $m$  sayıda değişken, Saaty (1977) tarafından önerilen karşılaştırma ölçeği kullanılarak birbirlerine göre önem değerlerine göre  $A$  karşılaştırma matrisi

$$A = [a_{ij}]_{m \times m} = \begin{bmatrix} 1 & a_{12} & \cdots & a_{1m} \\ 1/a_{12} & 1 & \cdots & a_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ 1/a_{1m} & 1/a_{2m} & \cdots & 1 \end{bmatrix} \quad (17)$$

biçiminde oluşturulur. Burada,  $a_{ij}$   $i$ . değişkenin  $j$ . değişkene göre önem değerini,  $1/a_{ij}$   $j$ . değişkenin  $i$ . değişkene göre önem değerini belirtir.

**Adım2:**  $A$  matrisi

$$a'_{ij} = \frac{a_{ij}}{\sum_{j=1}^m a_{ij}}, \quad i = 1, 2, \dots, m \quad (18)$$

biçiminde normleştirilir.

**Adım3:** Normleştirilmiş karşılaştırma matrisinin satır ortalamaları  $j$ . değişkene göre önem ağırlığı olup

$$w_j = \frac{1}{m} \sum_{i=1}^m a'_{ij}, \quad i = 1, 2, \dots, m \quad (19)$$

dır.

##### 3.1.2. Entropi yöntemi

Entropi yöntemi, değişken ağırlıklarının nesnel olarak belirlenmesinde kullanılan bir ÇKKV yöntemidir (Shannon, 1948). Entropi yöntemi uygulama adımları aşağıda verilmiştir.

**Adım1:**  $n$  gözlemlili bir veri setinde  $m$  sayıda bağımsız değişkenden oluşan bir  $X$  veri matrisi  $n \times m$  boyutlu

$$X = [x_{ij}]_{n \times m} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1m} \\ x_{21} & x_{22} & \cdots & x_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nm} \end{bmatrix} \quad (20)$$

olarak oluşturulur.

**Adım2:** Karar matrisi

$$x_{ij}^* = \frac{x_{ij}}{\sum_{j=1}^m x_{ij}}, \quad i = 1, 2, \dots, n \quad (21)$$

biçiminde normalleştirilir. Normalleştirilmiş karar matrisi  $N = [x_{ij}^*]_{n \times m}$  oluşturulur.

**Adım3:** Her bir değişken için Entropi değeri

$$E_j = -\frac{1}{\ln(n)} \sum_{i=1}^n x_{ij}^* \ln(x_{ij}^*) \quad j = 1, 2, \dots, m \quad (22)$$

olarak hesaplanır.

**Adım4:** Her bir değişken için belirsizlik değeri

$$d_j = 1 - E_j, \quad j = 1, 2, \dots, m \quad (23)$$

biçiminde hesaplanır.

**Adım5:** Değişken ağırlıkları

$$w_j = \frac{d_j}{\sum_{j=1}^m d_j}, \quad j = 1, 2, \dots, m \quad (24)$$

biçiminde belirlenir.

### 3.1.3. CRITIC yöntemi

CRITIC yöntemi, [Diakoulaki ve ark. \(1995\)](#) tarafından geliştirilen nesnel ağırlıklandırmada kullanılan ÇKKV yöntemlerinden biridir.

**Adım1:**  $X$  karar matrisi Eşitlik (20)'deki gibi tanımlanır.

**Adım2:** Karar matrisi

$$x_{ij}^* = \frac{x_{ij}}{\max x_{ij}}, \quad i = 1, 2, \dots, n, \quad j = 1, 2, \dots, m \quad (25)$$

biçiminde normalleştirilerek normalleştirilmiş karar matrisi  $N = [x_{ij}^*]_{n \times m}$  oluşturulur.

**Adım3:**  $s_j$ ,  $j$ . değişkenin standart sapması ve  $r_{jt}$ ,  $j$ . ve  $t$ . değişkenler arasındaki korelasyon katsayısı, sırasıyla

$$s_j = \sqrt{\frac{\sum_{i=1}^n (x_{ij}^* - \bar{x}_j^*)^2}{n}}$$

ve

$$r_{jt} = \frac{\sum_{i=1}^n (x_{ij}^* - \bar{x}_j^*)(x_{it}^* - \bar{x}_t^*)}{\sqrt{\sum_{i=1}^n (x_{ij}^* - \bar{x}_j^*)^2 \sum_{i=1}^n (x_{it}^* - \bar{x}_t^*)^2}}, \quad j, t = 1, 2, \dots, m \quad (26)$$

eşitlikleri ile hesaplanır.

**Adım4:** Her bir değişkene özgü bilgi miktarı

$$C_j = s_j \sum_{t=1}^m (1 - r_{jt}), \quad j = 1, 2, \dots, m \quad (27)$$

olarak hesaplanır.

**Adım5:** Değişken ağırlıkları

$$w_j = \frac{C_j}{\sum_{j=1}^m C_j}, \quad j = 1, 2, \dots, m \quad (28)$$

biçiminde belirlenir.

### 3.1.4. IDOCRIW yöntemi

IDOCRIW (Integrated Determination of Objective CRiteria Weights) yöntemi (Zavadskas & Podvezko, 2016), Entropi ve CILOS yöntemlerinin güçlü yönlerini birleştiren nesnel bir ağırlıklandırma yaklaşımıdır. Farklı ÇKKV yöntemleri ile hesaplanan değişkenlerin önem ağırlıklarının belirlenmesi amacıyla uygulanır.

**Adım1:**  $X$  karar matrisi Eşitlik (20)'deki gibi tanımlanır.

**Adım2:** Karar matrisi

$$x_{ij}^* = \frac{x_{ij}}{\sum_{j=1}^m x_{ij}}, \quad i = 1, 2, \dots, n \quad (29)$$

normalleştirilerek normalleştirilmiş karar matrisi  $N = [x_{ij}^*]_{n \times m}$  oluşturulur.

**Adım3:** Entropi ağırlıkları Eşitlikler (22-24) kullanılarak hesaplanır.

**Adım4:** Maliyet özelliğine sahip kriterler fayda özelliğine sahip kriterlere dönüştürülür. Fayda özelliği taşıyan kriterlerde bu dönüşüm yapılmadığından normalleştirilmiş karar matrisinde her bir sütunun en büyük değeri belirlenir. Elde edilen  $A_j = [a_{ij}]$  matrisinde köşegen elemanları  $a_{jj}$  olarak tanımlanır.

**Adım5:** Göreceli etki kaybı

$$p_{ij} = \frac{a_{jj} - a_{ij}}{a_{jj}} \quad (30)$$

hesaplanır.

**Adım6:**  $P$  matrisinin sütunları toplanarak  $S_j$  hesaplanır.  $f_{jj} = -S_j$  ve  $f_{ij} = p_{ij}$  olmak üzere  $F$  ağırlık sistemi matrisi hesaplanır.

**Adım7:** CILOS ağırlıkları hesaplanır. Denklem

$$F_w = 0 \sum_{j=1}^m w_{ij}^{(2)} = 1 \quad (31)$$

çözülerek  $w_{ij}^{(2)}$  CILOS ağırlıkları elde edilir.

**Adım8:** Nihai IDOCRIW ağırlıkları elde edilir.

$$w_j^* = \frac{w_j w_j^{(2)}}{\sum_{j=1}^m w_j w_j^{(2)}} \quad (32)$$

Hem bilgi teorisine hem etkileşim analizine dayalı nesnel ağırlıklar elde edilir.

### 3.2. Sıralama

Makine öğrenmesi sınıflandırma algoritmalarının performans ölçütlerine göre değerlendirilmesi amacıyla ÇKKV yöntemleri kullanılır. Sınıflandırma performans sonuçları, bir karar matrisi çerçevesinde ele alınarak, ÇKKV yöntemleri aracılığıyla karar süreci daha sistematik bir yapıya dönüştürülür.  $D$  karar matrisi

$$D = [d_{ij}]_{n \times m} = \begin{bmatrix} d_{11} & d_{12} & \cdots & d_{1m} \\ d_{21} & d_{22} & \cdots & d_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ d_{n1} & d_{n2} & \cdots & d_{nm} \end{bmatrix} \quad (33)$$

tanımlanır.  $D$  karar matrisi her bir ÇKKV yöntemi için Çizelge 4'teki uygun formül kullanılarak

$$V = [v_{ij}]_{n \times m} = \begin{bmatrix} v_{11} & v_{12} & \cdots & v_{1m} \\ v_{21} & v_{22} & \cdots & v_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ v_{n1} & v_{n2} & \cdots & v_{nm} \end{bmatrix} \quad (34)$$

normalize edilir (Vafaei, 2022).

Çizelge 4. Normalize yaklaşımlarına ilişkin formüller

| Yaklaşımlar | Amacı | Formüller                                                                                                       |
|-------------|-------|-----------------------------------------------------------------------------------------------------------------|
| Toplam      | Fayda | $v_{ij} = \frac{d_{ij}}{\sqrt{\sum_{k=1}^n d_{kj}}} \left( \frac{d_{ij}}{\sqrt{\sum_{k=1}^n d_{kj}^2}} \right)$ |
| Oran        | Fayda | $v_{ij} = \frac{d_{ij}}{\max d_{ij}}$                                                                           |
| Min max     | Fayda | $v_{ij} = \frac{d_{ij} - \min_j(d_{ij})}{\max_j(d_{ij}) - \min_j(d_{ij})}$                                      |

ÇKKV yöntemlerine ilişkin uygulama adımları Çizelge 5’te verilmiştir (Hwang & Yoon, 1981; Zavadskas & Kaklauskas, 1996; Ghorabae ve ark., 2015; Ghorabae ve ark., 2016; Pamučar & Ćirović, 2015; Özçalıcı, 2017; Kabak & Çınar, 2020; Xie & Zhang 2023). Çizelge 5’te, *Adım1*’de yer alan *D* matrisi Eşitlik (33)’te ve *Adım2*’de yer alan *V* matrisi Eşitlik (34)’te tanımlandığı gibi hesaplanır.

Çizelge 5. ÇKKV Yöntemleri ile alternatiflerin sıralama adımları (  $i, k = 1, 2, \dots, n$ ,  $j = 1, 2, \dots, m$  )

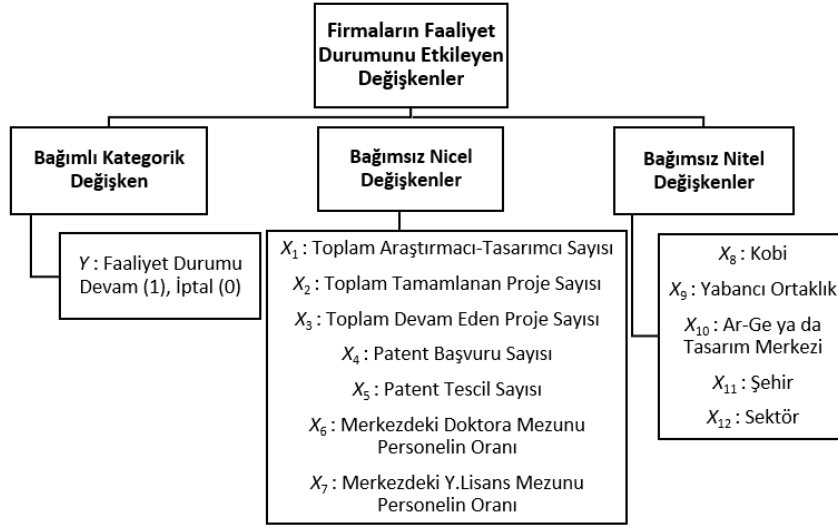
| Yöntem        | Adım 1   | Adım 2   | Adım 3                                                                                                        | Adım 4                                                                                                                                                                                                                                                                                                                                                                                                                                                        | Adım 5                              |
|---------------|----------|----------|---------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------|
| <b>TOPSIS</b> | <i>D</i> | <i>V</i> | $V^+ = [v_j^+]_{1 \times m}$ $V^- = [v_j^-]_{1 \times m}$                                                     | $S_i^+ = \sqrt{\sum_{j=1}^m (v_{ij} - v_j^+)^2}$ $S_i^- = \sqrt{\sum_{j=1}^m (v_{ij} - v_j^-)^2}$                                                                                                                                                                                                                                                                                                                                                             | $C_i = \frac{S_i^-}{S_i^+ + S_i^-}$ |
| <b>CODAS</b>  | <i>D</i> | <i>V</i> | $V^- = [v_j^-]_{1 \times m}$                                                                                  | $E_i = \sqrt{\sum_{j=1}^m (v_{ij} - v_j^-)^2}$ $T_i = \sum_{j=1}^m  v_{ij} - v_j^- $ $h_{ik} = (E_i - E_k) + \psi(E_i - E_k) \times (T_i - T_k)$                                                                                                                                                                                                                                                                                                              | $H_i = \sum_{k=1}^n h_{ik}$         |
| <b>EDAS</b>   | <i>D</i> | <i>V</i> | $PDA_{ij} = \frac{\max(0, (d_{ij} - v_{ij}))}{v_{ij}}$ $NDA_{ij} = \frac{\max(0, (v_{ij} - d_{ij}))}{v_{ij}}$ | $SP_i = \sum_{j=1}^m PDA_{ij}$ $SN_i = \sum_{j=1}^m NDA_{ij}$ $NSP_i = \frac{SP_i}{\max_i(SP_i)}$ $NSN_i = 1 - \frac{SN_i}{\max_i(SN_i)}$                                                                                                                                                                                                                                                                                                                     | $AS_i = \frac{NSP_i + NSN_i}{2}$    |
| <b>MABAC</b>  | <i>D</i> | <i>V</i> | $g_j = (\prod_{i=1}^n v_{ij})^{(1/n)}$ $G = [g_1 \ g_2 \ \dots \ g_m]$                                        | $Q = [q_{ij}]_{n \times m} = V - G$ $Q = \begin{bmatrix} v_{11} - g_1 & v_{12} - g_2 & \dots & v_{1m} - g_m \\ v_{21} - g_1 & v_{22} - g_2 & \dots & v_{2m} - g_m \\ \vdots & \vdots & \ddots & \vdots \\ v_{n1} - g_1 & v_{n2} - g_2 & \dots & v_{nm} - g_m \end{bmatrix}$ $Q = \begin{bmatrix} q_{11} & q_{12} & \dots & q_{1m} \\ q_{21} & q_{22} & \dots & q_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ q_{n1} & q_{n2} & \dots & q_{nm} \end{bmatrix}$ | $S_i = \sum_{j=1}^m q_{ij}$         |

#### 4. UYGULAMA

Bu çalışmada, Ar-Ge teşviklerinden yararlanmak amacıyla Ar-Ge ve Tasarım merkezi belgesine sahip firmaların faaliyetlerinin devam ettirilmesi ya da iptal edilmesi (belge iptali) durumlarına göre sınıflandırma çalışması yapılması istenmiştir. Veri seti, 2008-2021 yılları arasında faaliyeti devam eden Ar-Ge ve Tasarım Merkezleri'ni kapsamaktadır.

##### Veri Ön İşleme

Çalışma kapsamında, bilgi kaybının istenmemesi nedeniyle verinin tamamı değerlendirilmiş olup ilgilenilen veri setine, öncelikli olarak veri ön işleme uygulanmıştır. Benzer bilgiyi içeren değişkenler birleştirilerek daha anlamlı hale getirilmiş, değişken seçimi ve boyut indirgeme yapılmıştır. Çalışmada yapılan analizler için Python 3.11.3, PyDecision kütüphanesi kullanılmıştır. İlgilenilen veri setine yönelik, bağımlı ve bağımsız değişkenler belirlenerek, değişkenlerin aldığı değerlere ilişkin kategorileri Şekil 1'de detaylı biçimde belirtilmiştir (Ulu Metin, 2024).

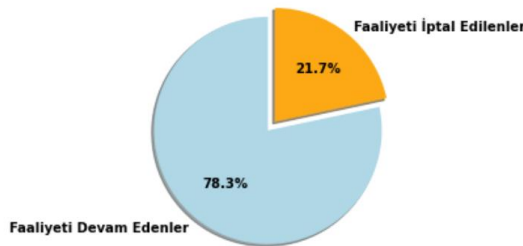


Şekil 1. Firmaların faaliyet durumunu etkileyen değişkenler.

Şekil 1'de, veri seti yedi nicel, beş nitel değişken ve bir kategorik bağımlı değişken, olmak üzere toplam on iki (12) bağımsız değişkenden oluşmaktadır. Veri setinde, veri temizliği sonrası 2334 merkez verisi kullanılmıştır.

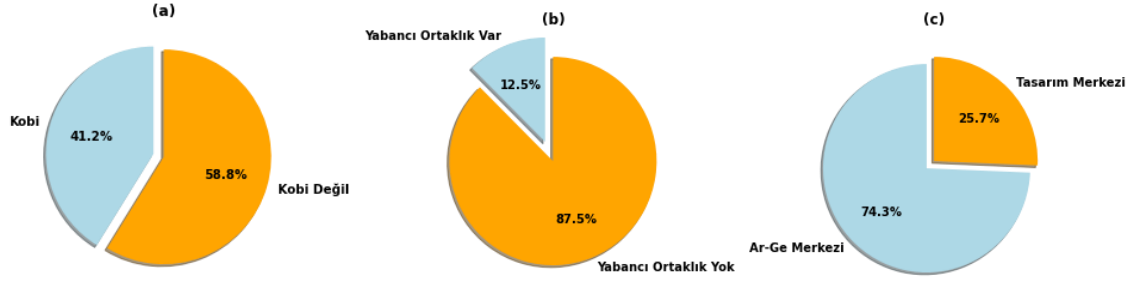
##### Betimsel İstatistikler ve Veri Görselleştirme

Değişkenlere ait betimsel istatistikler ve ilgili veri görselleri, veri seti hakkında genel bilgi sahibi olabilmek amacıyla verilmiştir.



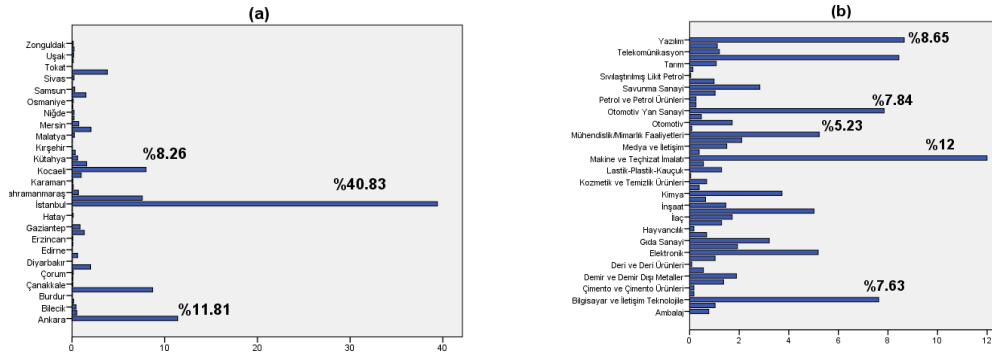
Şekil 2. Y bağımlı değişkenine ait daire grafiği.

Şekil 2’de, Ar-Ge ve Tasarım merkezlerinin yaklaşık %78.3’ü faaliyette olup %21.7’sinin ise faaliyetleri devam etmemektedir.



Şekil 3. (a)  $X_8$ , (b)  $X_9$  ve (c)  $X_{10}$  bağımsız değişkenlerine ait daire grafiği.

Şekil 3 ve Şekil 4’te veri setindeki nitel değişkenlere ait grafikler gösterilmiştir. Şekil 3(a)’da, Ar-Ge ve Tasarım merkezlerinin yaklaşık %59’u büyük ölçekli işletme sınıfında yer alırken %41’i Kobi’dir. Şekil 3(b)’de Ar-Ge ve Tasarım merkezlerinin yaklaşık %87’sinin yabancı ortaklığı bulunmazken %12’sinin yabancı ortaklığı bulunmaktadır. Şekil 3(c)’de ise firmaların yaklaşık %74’ü Ar-Ge merkezi iken %26’sı Tasarım merkezidir.



Şekil 4. (a)  $X_{11}$  ve (b)  $X_{12}$  bağımsız değişkenlerine göre yüzdelik değerler.

Şekil 4(a)’da, Ar-Ge veya Tasarım merkezlerinin %40.83’ü İstanbul’da, %11.81’i Ankara’da, %8.26’sı Kocaeli’nde bulunmaktadır. Şekil 4(b)’de, Ar-Ge ve Tasarım merkezlerinin %12’si makine imalatı, %8.65’i yazılım, %7.84’ü ise otomotiv yan sanayi ve %5.23’ü mühendislik-mimarlık sektöründe faaliyet gösterdiği görülmektedir.

Çizelge 6. Betimsel istatistikler (Nicel değişkenler)

|       | $X_1$ | $X_2$ | $X_3$ | $X_4$ | $X_5$ | $X_6$ | $X_7$ |
|-------|-------|-------|-------|-------|-------|-------|-------|
| $n$   | 2334  | 2334  | 2334  | 2334  | 2334  | 2334  | 2334  |
| $min$ | 1     | 0     | 0     | 0     | 0     | 0     | 0     |
| $Q_1$ | 11    | 5     | 4     | 0     | 0     | 0     | 0.030 |
| $Q_2$ | 15    | 12    | 6     | 0     | 0     | 0     | 0.080 |
| $Q_3$ | 24    | 26    | 11    | 2     | 1     | 0     | 0.140 |
| $max$ | 2710  | 1056  | 485   | 2127  | 693   | 0.43  | 0.710 |

Çizelge 6’da Ar-Ge ya da Tasarım merkezi olan firmaların yarısında 15’ten fazla Araştırmacı/Tasarımcı personelin bulunduğu ve bu sayının en fazla 2710 olduğu, toplam tamamlanan ve toplam devam eden proje sayısı sıfır olan firmaların faaliyette olmadığı bu nedenle sayılarının sıfır olarak görüldüğü, en fazla 2127 patent başvurusu yapıldığı, en fazla 693 patente tescil alındığı, en fazla doktoralı oranının % 43, en fazla yüksek lisanslı oranının % 71 olduğu görülmektedir (Ulu Metin & Türkşen, 2023).

#### Çok Kriterli Karar Verme ile Ağırlıklandırma

Ar-Ge ve Tasarım Merkezleri Dairesi’nde çalışan uzmanların görüşlerini dikkate almak amacıyla nitel ve nicel değişkenleri değerlendirebilen öznel ağırlıklandırma yönteminden AHP yöntemi seçilmiştir. AHP yönteminde her bir değişkenin birbirine göre doğrudan öneminin belirlenebilmesi amacıyla Ar-Ge ve Tasarım Merkezleri Dairesi’nde çalışan ve en az 5 yıl Ar-Ge projelerini değerlendirme deneyimine sahip 15 kişi (Bilgisayar Mühendisi, Endüstri Mühendisi, Makine Mühendisi, Tekstil Mühendisi, İstatistik, Kamu Yönetimi ve İktisat bölümü lisans mezunları) olmak üzere farklı disiplinlerden seçilmiştir. Uzmanların görüşü alınarak karşılaştırma matrisi oluşturulmuş olup AHP ile değişken ağırlıkları Eşitlik (19) kullanılarak belirlenmiştir.

Çizelge 7. AHP ile belirlenen değişken ağırlıkları

|              | <i>Değişkenler</i> | <i>w</i> |
|--------------|--------------------|----------|
| <b>Nicel</b> | $X_1$              | 0.131    |
|              | $X_2$              | 0.102    |
|              | $X_3$              | 0.102    |
|              | $X_4$              | 0.080    |
|              | $X_5$              | 0.121    |
|              | $X_6$              | 0.183    |
|              | $X_7$              | 0.118    |
| <b>Nitel</b> | $X_8$              | 0.037    |
|              | $X_9$              | 0.024    |
|              | $X_{10}$           | 0.018    |
|              | $X_{11}$           | 0.040    |
|              | $X_{12}$           | 0.039    |

AHP ile belirlenen değişken ağırlıkları Çizelge 7’de verilmiştir. Buna göre, merkezlerin faaliyetini etkileyen başlıca değişkenlerin  $X_6$  (Merkezdeki Doktora Mezunu Personelin Oranı) ve  $X_1$  (Araştırmacı/Tasarımcı Sayısı)’dir. Merkeze ait Sektör, Şehir, Kobi, Ar-Ge/Tasarım Merkezi ve Yabancı Ortaklık bilgilerinin Ar-Ge ve Tasarım merkezlerinin faaliyet durumunu etkilemede en az öneme sahip olduğu görülmektedir. AHP ile belirlenen nitel değişken ağırlıklarının düşük olması sebebiyle nicel değişkenleri kullanarak nesnel ağırlık belirleyen ÇKKV yöntemlerinden Entropi ve CRITIC yöntemleri seçilmiştir. Eşitlik (24) ve Eşitlik (28) ile değişken ağırlıkları hesaplanmıştır. Hesaplanan nesnel değişken ağırlıkları Çizelge 8’de verilmiştir.

Çizelge 8. Entropi ve CRITIC yöntemleri ile belirlenen değişken ağırlıkları

| Değişkenler | Entropi | CRITIC |
|-------------|---------|--------|
| $X_1$       | 0.084   | 0.360  |
| $X_2$       | 0.076   | 0.197  |
| $X_3$       | 0.044   | 0.064  |
| $X_4$       | 0.285   | 0.269  |
| $X_5$       | 0.280   | 0.107  |
| $X_6$       | 0.186   | 0.0001 |
| $X_7$       | 0.042   | 0.0004 |

#### Makine Öğrenmesi Sınıflandırma Algoritmaları

LR, kNN, SVM ve RF makine öğrenmesi sınıflandırma algoritmaları ağırlıklandırılmış veri setine uygulanmıştır. Sınıflandırma algoritmalarının performansının ölçülebilmesi amacıyla öncelikle veri setinin %80'i eğitim seti, %20'si test seti olmak üzere ayrılmıştır. Sınıflandırma algoritmalarının performansı 5-kat çapraz doğrulama kullanılarak test edilmiş, optimal parametreler ızgara arama yöntemi ile elde edilmiştir.

Çizelge 9. Izgara arama ile belirlenen algoritma ayarlanabilir parametreleri

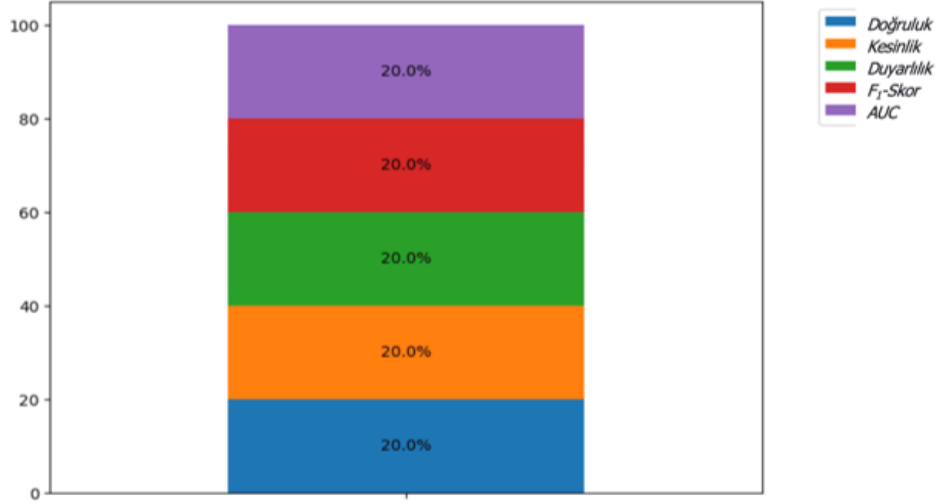
| Sınıflandırma Algoritmaları | Ayarlanabilir Parametreleri                                                                                                                                                     |
|-----------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| LR                          | $Penalty = \text{None}$ , $c = 1e-06$                                                                                                                                           |
| kNN                         | $Neighbors(k) = 5$ , $Metric = \text{Euclidean}$                                                                                                                                |
| SVM                         | $Kernel = \text{Rbf}$ , $Gamma = 10$ , $c = 100$                                                                                                                                |
| RF                          | $Min\ Samples\ Leaf = 5$ , $Min\ Samples\ Split = 3$ , $Criterion = \text{Gini}$<br>$n\ Estimator = 350$ , $Bootstrap = \text{True}$ , $Max\ Depth = 5$ , $Max\ Features = 3$ , |

Her bir yöntem için kullanılan ayarlanabilir parametreler Çizelge 9'da verilmiştir. AHP, Entropi ve CRITIC yöntemleri ile ağırlıklandırılmış veri setine LR, kNN, SVM ve RF makine öğrenmesi sınıflandırma algoritmalarının uygulanmasıyla elde edilen yöntemlere ilişkin performans değerleri Çizelge 10'da verilmiştir.

Çizelge 10. Ağırlıklandırılmış veri setine ilişkin performans değerleri (AHP, Entropi ve CRITIC)

| Yöntemler   | Doğruluk | Kesinlik | Duyarlılık | F <sub>1</sub> -Skor | AUC    |
|-------------|----------|----------|------------|----------------------|--------|
| AHP_LR      | 0.8522   | 0.8906   | 0.9226     | 0.9063               | 0.9077 |
| AHP_kNN     | 0.8308   | 0.8713   | 0.9171     | 0.8936               | 0.8540 |
| AHP_SVM     | 0.8522   | 0.9013   | 0.9088     | 0.9050               | 0.8936 |
| AHP_RF      | 0.8650   | 0.8986   | 0.9309     | 0.9145               | 0.9140 |
| Entropi_LR  | 0.8543   | 0.8909   | 0.9254     | 0.9078               | 0.9083 |
| Entropi_kNN | 0.8458   | 0.9073   | 0.8922     | 0.8997               | 0.8711 |
| Entropi_SVM | 0.8608   | 0.9002   | 0.9226     | 0.9113               | 0.9101 |
| Entropi_RF  | 0.8650   | 0.8986   | 0.9309     | 0.9145               | 0.9140 |
| CRITIC_LR   | 0.8458   | 0.8918   | 0.9116     | 0.9016               | 0.9056 |
| CRITIC_kNN  | 0.8286   | 0.8790   | 0.9033     | 0.8910               | 0.8849 |
| CRITIC_SVM  | 0.8586   | 0.9088   | 0.9088     | 0.9088               | 0.8987 |
| CRITIC_RF   | 0.8650   | 0.8986   | 0.9309     | 0.9145               | 0.9140 |

AHP, Entropi ve CRITIC ağırlıkları ile hesaplanan sınıflandırma performans ölçütlerinin önem ağırlıkları hesaplanmak istenmiştir. Buna göre, satırda sınıflandırma yöntemleri (AHP\_LR, AHP\_kNN,...,CRITIC\_RF), sütunda performans değerleri (Doğruluk, Kesinlik, Duyarlılık,  $F_1$ -Skor ve AUC) olmak üzere  $(12 \times 5)$  boyutlu analiz sonuçları bir karar matrisi olarak ele alınarak IDOCRIW ile performans ölçütlerinin ağırlıkları elde edilmiştir. Performans ölçütlerinin eşit önem ağırlıklarına sahip olduğu Şekil 5'te görülmektedir.



Şekil 5. IDOCRIW performans ölçüt ağırlıkları.

#### Çok Kriterli Karar Verme ile Sıralama

Performans değerleri bakımından sınıflandırma algoritmalarına karar verilmesi bir ÇKKV problemi olduğundan, objektif karar verebilmek için performans değerleri bir karar matrisi olarak ele alınıp performans değerleri maksimum (fayda) olacak biçimde TOPSIS, EDAS, MABAC ve CODAS yöntemleri Çizelge 5'te verilen adımlar uygulanarak elde edilmiştir. ÇKKV skorları Çizelge 11'de verilmiştir.

Çizelge 11. ÇKKV skorları

| Yöntemler   | TOPSIS | EDAS  | MABAC  | CODAS  |
|-------------|--------|-------|--------|--------|
| AHP_LR      | 0.684  | 0.622 | 0.078  | 0.014  |
| AHP_kNN     | 0.252  | 0.000 | -0.458 | -0.107 |
| AHP_SVM     | 0.619  | 0.496 | 0.005  | -0.015 |
| AHP_RF      | 0.887  | 1.000 | 0.324  | 0.060  |
| Entropi_LR  | 0.721  | 0.686 | 0.120  | 0.021  |
| Entropi_kNN | 0.441  | 0.300 | -0.204 | -0.054 |
| Entropi SVM | 0.838  | 0.838 | 0.227  | 0.039  |
| Entropi_RF  | 0.887  | 1.000 | 0.324  | 0.060  |
| CRITIC_LR   | 0.561  | 0.475 | -0.055 | -0.008 |
| CRITIC_kNN  | 0.254  | 0.099 | -0.420 | -0.084 |
| CRITIC_SVM  | 0.714  | 0.682 | 0.130  | 0.014  |
| CRITIC_RF   | 0.887  | 1.000 | 0.324  | 0.060  |

Çizelge 11'de verilen ÇKKV skorlarına göre oluşturulan ÇKKV sıralamaları Çizelge 12'de verilmiştir.

Çizelge 12. ÇKKV sıralamaları

| Yöntemler   | TOPSIS | EDAS | MABAC | CODAS |
|-------------|--------|------|-------|-------|
| AHP_RF      | 2      | 2    | 2     | 2     |
| Entropi_RF  | 2      | 2    | 2     | 2     |
| CRITIC_RF   | 2      | 2    | 2     | 2     |
| Entropi_SVM | 4      | 4    | 4     | 4     |
| Entropi_LR  | 5      | 5    | 6     | 5     |
| CRITIC_SVM  | 6      | 6    | 5     | 6     |
| AHP_LR      | 7      | 7    | 7     | 7     |
| AHP_SVM     | 8      | 8    | 8     | 9     |
| CRITIC_LR   | 9      | 9    | 9     | 8     |
| Entropi_kNN | 10     | 10   | 10    | 10    |
| CRITIC_kNN  | 11     | 11   | 11    | 11    |
| AHP_kNN     | 12     | 12   | 12    | 12    |

Çizelge 12’de sınıflandırma algoritmalarının ÇKKV yöntemlerine göre tercih sıralamaları verilmiştir. Buna göre, AHP, Entropi ve CRITIC ağırlıklandırma yöntemlerinin her biri kullanılarak yapılan sınıflandırmada RF algoritmasının öncelikli olarak tercih edildiği açıkça görülmektedir. kNN algoritmasının ise son sırada tercih edilebilir olduğu sonucuna ulaşılmıştır.

## 5. SONUÇ

Çalışmada Ar-Ge teşvikleri alanındaki bir sınıflandırma problemi ele alınmıştır. Bu amaçla, Ar-Ge ve Tasarım merkezlerine ait gerçek bir veri seti uygulama amacıyla kullanılmıştır. Farklı önem derecelerine sahip olduğu düşünülen veri setindeki değişkenler hem öznel (AHP) hem de nesnel yaklaşım (Entropi ve CRITIC) kullanılarak ağırlıklandırmıştır. Gerçek veri setinin amacına uygun bir biçimde uygulanan makine öğrenmesi sınıflandırma algoritmalarının (LR, kNN, SVM ve RF) performans değerleri (Doğruluk, Duyarlılık, Kesinlik, F<sub>1</sub>-Skor ve AUC) 5-kat çapraz doğrulama ile hesaplanmıştır. Sınıflandırma performans ölçütlerinin kriter ağırlıklarının hesaplanmasında IDOCRIW yöntemi kullanılmıştır. Farklı ÇKKV yöntemleri ile ağırlıklandırılmış sınıflandırma algoritmalarının (AHP\_LR, AHP\_kNN, CRITIC\_RF, vb.) tercih sıralaması, ÇKKV yöntemleri (TOPSIS, EDAS, MABAC ve CODAS) ile belirlenmiştir. Elde edilen sonuçlara göre farklı ağırlıklandırma yöntemleri ile uygulanan sıralama yöntemlerinin sınıflandırma performans değerleri bakımından RF algoritmasının (AHP\_RF, Entropi\_RF ve CRITIC\_RF) öncelikli olarak tercih edilebileceğini açıkça göstermektedir. Bu yönüyle çalışma, Ar-Ge teşviklerinin değerlendirilmesinde RF sınıflandırma algoritmasının karar vericiler için tercih edilebilir olduğunu vurgulamaktadır. Böylelikle bu yaklaşım, Ar-Ge teşvik politikalarının şeffaflığını, esnekliğini ve veriye dayalı yapısını güçlendirerek Ar-Ge ekosistemine önemli bir katkı sağlamaktadır.

## Teşekkür / Destekleyen Kuruluş

Çalışmanın uygulama kısmında kullanılan veri setinin temin edilmesindeki katkılarından dolayı Sanayi ve Teknoloji Bakanlığı Ar-Ge Teşvikleri Genel Müdürlüğü Ar-Ge ve Tasarım Dairesi Başkanlığı’na teşekkür ederiz.

## Kongre ve/veya Tez Beyanı

Makalede yer verilen uygulamaların bir kısmı 14-16 Mayıs 2025 tarihlerinde Ankara’da düzenlenen VI. Uluslararası Uygulamalı İstatistik Kongresi (UYİK2025)’te sözlü olarak sunulmuştur. Bu çalışma, birinci yazarın ikinci yazarın danışmanlığında hazırlayarak yayımladığı Doktora tezinden üretilmiştir.

## Katkı Oranı Beyanı

**Gözde Ulu Metin:** Araştırma, Metodoloji, Veri Analizi, Yazılım, Görselleştirme, Gözden Geçirme ve Düzenleme. **Özlem Türkşen:** Metodoloji, Taslak Yazımı, Gözden Geçirme ve Düzenleme, Biçimsel Analiz, Danışmanlık.

## Çıkar Çatışması Beyanı

Yazarlar, herhangi bir çıkar çatışması olmadığını beyan ederler.

## Araştırma ve Yayın Etiği Beyanı

Bu makalenin yazarlar çalışmalarında araştırma ve yayın etiğine uyduklarını beyan ederler:

## Etik Kurul Beyanı

Bu makalenin yazarlar çalışmada kullanılan materyal ve yöntemlerin etik kurul izni ve/veya yasal-özel bir izin gerektirmediğini beyan ederler.

## Yapay Zeka Kullanımı

Yazarlar, bu makalenin yazımında, görsellerin, grafiklerin, tabloların ya da bunlara karşılık gelen başlıkların oluşturulmasında herhangi bir tür üretken yapay zeka kullanmadıklarını beyan ederler.

## Kaynakça

- Arılı, N. B., Gürsakil S., & Engin, M. (2022). Denetimli makine öğrenmesi algoritmaları R ve Python uygulamaları. *Nobel Akademik Yayıncılık*, 294 s., Ankara.
- Diakoulaki, D., Mavrotas, G., & Papayannakis, L. (1995). Determining objective weights in multiple criteria problems: the critic method. *Computers & Operations Research*, 22(7), 763-770.
- Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). From data mining to knowledge discovery in databases. *AI Magazine*, 17(3), 37-37.
- Geron, A. (2017). Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow. *O'Reilly Media, Inc.*, 1005, 564.
- Geron, A. (2021). Scikit-Learn Keras ve TensorFlow ile Uygulamalı Makine Öğrenmesi. Çeviri. *Buzdağı Yayınevi*, 862 s., Ankara.
- Ghorabae, M. K., Zavadskas, E. K., Olfat, L., & Turskis, Z. (2015). Multi-Criteria inventory classification using a new method of evaluation based on distance from average solution (EDAS). *Informatica*, 26(3), 435-451. <https://doi.org/10.3233/INF-2015-10>
- Ghorabae, M. K., Zavadskas, E. K., Turskis, Z., & Antucheviciene, J. (2016). A new combinative distance-based assessment (CODAS) method for multi-criteria decision-making. *Economic Computation and Economic Cybernetics Studies and Research*, 50, 25-44.
- Hossin, M., & Sulaiman, M. N. (2015). A review on evaluation metrics for data classification evaluations. *International Journal of Data Mining & Knowledge Management Process*, 5, 01-11.
- Hwang, C. L., & Yoon, K. (1981). Multiple Attribute Decision Making: Methods and Applications. *Springer- Verlag*, 228 p., New York.
- Joshi, M. V. (2002). On Evaluating Performance of Classifiers for Rare Classes. *IEEE International Conference on Data Mining. Proceedings*, 641-644, IEEE. <https://doi.org/10.1109/ICDM.2002.1184018>
- Kabak, M., & Çınar, Y. (2020). Çok Kriterli Karar Verme Yöntemleri: Ms Excel ® Çözümlü Uygulamalar. *Nobel Akademik Yayıncılık*, 127 s., Ankara.
- Kumar, A., Singh, S. S., Singh, K., & Biswas, B. (2020). Link Prediction Techniques, Applications, and Performance: A Survey. *Physica A: Statistical Mechanics and its Applications*, 553, 124289.

- Metlek, S., & Kayaalp, K. (2020). Makine Öğrenmesinde, Teoriden Örnek MATLAB Uygulamalarına Kadar Destek Vektör Makineleri. *İksad Yayınevi*, 98 s., Ankara.
- Özçalıcı, M. (2017). Matlab ile Çok Kriterli Karar Verme Teknikleri. Nobel Akademik Yayıncılık, 566 s., Ankara.
- Pamuçar, D., & Ćirović, G. (2015). The Selection of Transport and Handling Resources in Logistics Centers Using Multi-Attributive Border Approximation Area Comparison (MABAC). *Expert Systems with Applications*, 42(6), 3016-3028.
- Rafiei-Sardooi, E., Azareh, A., Choubin, B., Mosavi, A. H. & Clague, J. J. (2021). Evaluating Urban Flood Risk Using Hybrid Method of TOPSIS and Machine Learning. *International Journal of Disaster Risk Reduction*, 66, 102614.
- Saaty, T. L. (1977). A Scaling Method for Priorities in Hierarchical Structures. *Journal of Mathematical Psychology*, 15, 234-281.
- Shannon, C. E. (1948). A mathematical theory of communication. *The Bell System Technical Journal*, 27(3), 379-423. <https://doi.org/10.1002/j.1538-7305.1948.tb01338.x>
- Sorhun, E. (2021). Python ile Makine Öğrenmesi. 464 s., *Abaküs Kitap*, İstanbul.
- Türkşen, Ö. (2023). Optimizasyon Yöntemleri ve Matlab, Python, R Uygulamaları. *Nobel Akademik Yayıncılık*, 468 s., Ankara.
- Uğuz, S. (2021). Makine Öğrenmesi Teorik Yönleri ve Python Uygulamaları ile Yapay Zekâ Ekolü. *Nobel Akademik Yayıncılık*, 298 s., Ankara.
- Ulu Metin, G., & Türkşen, Ö. (2023). Veri madenciliği yöntemleri ile bir melez sınıflandırma yaklaşımı ve uygulaması. *İstatistikçiler Dergisi: İstatistik ve Aktüerya*, 16(2), 100–115.
- Ulu Metin, G. (2024). Ar-Ge ve tasarım merkezlerinin istatistiksel olarak değerlendirilmesi, veri madenciliği yöntemleri ile hibrit karar verme. (PhD), Ankara Üniversitesi, Fen Bilimleri Enstitüsü, Ankara, Türkiye.
- Vafaei, N., Ribeiro, R. A., & Camarinha-Matos, L. M. (2022). Assessing normalization techniques for simple additive weighting method. *Procedia Computer Science*, 199, 1229-1236. <https://doi.org/10.1016/j.procs.2022.01.156>
- Xie, S., & Zhang, J. (2023). TOPSIS-based comprehensive measure of variable importance in predictive modelling. *Expert Systems with Applications*, 232 (2023). <https://doi.org/10.1016/j.eswa.2023.120682>
- Zavadskas, E. K., & Kaklauskas, A. (1996). Systemotechnical Evaluation of Buildings (Pastatų sistemotechninis įvertinimas). *Vilnius: Technika*, 280 p. (in Lithuanian).
- Zavadskas, E. K., & Podvezko, V. (2016). Integrated determination of objective criteria weights in MCDM. *International Journal of Information Technology & Decision Making*, 15(02), 267-283.