



INTRODUCTION TO ARTIFICIAL INTELLIGENCE DERSİ

ARAŞTIRMA ÖDEVİ

FATİH TALHA TÜMER – 191180081

## İÇİNDEKİLER

|  |    |
|--|----|
| İÇİNDEKİLER.....                                 | i  |
| ŞEKİLLER LİSTESİ.....                            | ii |
| 1. GİRİŞ .....                                   | 1  |
| 2 SINIFLANDIRICILARIN PERFORMANS ÖLÇÜTLERİ ..... | 2  |
| 3 EĞİTİM VE TEST VERİLERİNİN SEÇİLMESİ .....     | 6  |
| 4 SONUÇ .....                                    | 8  |
| KAYNAKÇA .....                                   | 9  |

**ŞEKİLLER LİSTESİ**

|   |   |
|---|---|
| Şekil 1: Karmaşıklık matrisi.....             | 2 |
| Şekil 2: AUC-ROC eğrisi .....                 | 5 |
| Şekil 3: Rastgele örnekleme örneği.....       | 6 |
| Şekil 4: Katmanlı örnekleme örneği.....       | 6 |
| Şekil 5: Cross validation örneği .....        | 7 |
| Şekil 5: K-fold Cross-validation örneği ..... | 7 |

## 1. GİRİŞ

Bu çalışmada, sınıflandırıcıların performansını hesaplamada kullanılan metrikler araştırılmış; avantajları, dezavantajları ve kullanım alanları açıklanmıştır. Sonraki bölümde ise sınıflandırma öncesinde veri setinin eğitim/test olarak ayrılmasında kullanılan yöntemler açıklanmıştır. Sonuç bölümünde ise araştırmadan elde edilen kazanımlara yer verilmiştir.

## 2 SINIFLANDIRICILARIN PERFORMANS ÖLÇÜTLERİ

Sınıflandırma kavramı, verileri bir veri kümesinde tanımlanan çeşitli sınıflar arasında dağıtmaktır. Sınıflandırma algoritmaları verilen eğitim setinden bu dağılım örüntüsünü öğrenir ve daha sonra sınıfı belirlenmemiş test verileri alındığında doğru sınıflandırma yapmaya çalışır. Sınıflandırma algoritmalarına örnek olarak “Support Vector Machine”, “Naive Bayes Classifier” ve “Decision Tree” algoritmaları verilebilir [1].

Sınıflandırıcıların performansı, sınıflandırıcıların tahminlerinin doğruluğunun ve başarısının bir ölçüsüdür. Sınıflandırıcıların performansını ölçmek için birçok farklı metrik kullanılabilir. Hangi kriterlerin kullanılacağı uygulamanın amacına ve veri setinin özelliklerine göre değişiklik gösterebilir.

Sınıflandırıcıların performansını ölçmek için, sınıflandırıcının tahminlerinin gerçek sınıflarla karşılaştırılması gerekir. Bu, eğitim veri setinde zaten bilinen sınıflar kullanılarak eğitildiğinden, test veri seti genellikle sınıflandırıcının performansını ölçmek için kullanılır. Test veri setindeki örneklerin sınıfları önceden bilinmemekte ve sınıflandırıcının bu örnekleri doğru bir şekilde sınıflandırması beklenmektedir. Sınıflandırıcının tahminlerinin doğruluğu, test veri setindeki örneklerin sınıfları karşılaştırılarak hesaplanır. Bu noktada karmaşıklık matrisi (confusion matrix) kullanılır [2].

|                  |              | Actual Values |              |
|------------------|--------------|---------------|--------------|
|                  |              | Positive (1)  | Negative (0) |
| Predicted Values | Positive (1) | TP            | FP           |
|                  | Negative (0) | FN            | TN           |

Şekil 1: Karmaşıklık matrisi

Karmaşıklık matrisinde gerçek veriler doğru (true) ve yanlış (false), tahmin edilen veriler ise pozitif ve negatif olarak gruplandırılır.

- Actual values: Gerçek değerler

- Predicted values: Eğitilen model sonunda tahmin edilen değerler
- True positive: Pozitif tahmin edildi ve sonuç doğru
- True negative: Negatif tahmin edildi ve sonuç doğru
- False positive (Tip 1 hata): Pozitif tahmin edildi ve sonuç yanlış
- False negative (Tip 2 hata): Negatif tahmin edildi ve sonuç yanlış
- TP + FP = Modelin pozitif olarak tahmin ettiği veriler
- TN + FN = Modelin negatif olarak tahmin ettiği veriler
- TP + TN = Modelin doğru tahmin ettiği veriler
- FP + FN = Modelin yanlış tahmin ettiği veriler

Modelin performansı ölçülürken en çok kullanılan altı ölçüt accuracy (doğruluk), error rate (hata oranı), recall (duyarlılık), precision (kesinlik), f1-score, AUC-ROC ölçütleridir [2,3,4,5].

- Accuracy: Tüm tahminlerin doğruluğunu ölçmeye yarar. Genel performansı ölçmek için kullanılır ancak modelin performansını değerlendirmek için tek başına yeterli değildir.  
+ Basit olduğu için anlaşılması kolaydır.  
- Veri集中的indeki verilerin dengesizliğini yansıtmaz. Örneğin kanser ile ilgili bir veri setinde kanser olmayan verilerin oranı %90 bile olsa sınıflandırmanın accuracy değeri yüksek çıkabilir.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

Accuracy, genellikle geniş veri kümelerinde kullanılır ve sınıflar arasındaki dengesizliği yansıtmaz

- Error rate: Tüm tahminlerdeki hata oranını ölçmeye yarar. Genel performansı ölçmek için kullanılır ancak modelin performansını değerlendirmek için tek başına yeterli değildir.  
+ Basit olduğu için anlaşılması kolaydır.  
- Veri集中的indeki verilerin dengesizliğini yansıtmaz. Örneğin kanser ile ilgili bir veri setinde kanser olmayan verilerin oranı %90 bile olsa sınıflandırmanın error rates değeri düşük çıkabilir.

$$Error\ rate = 1 - Accuracy$$

Error rate, genellikle geniş veri kümelerinde kullanılır ve sınıflar arasındaki dengesizliği yansıtmaz

- Recall: Sınıflandırıcının pozitif olarak tahmin ettiği göstergelere sahip gerçek pozitiflerin oranıdır. Recall sınıflandırıcının pozitif sınıfı doğru olarak tahmin etme becerisini ölçer.
  - + Pozitif etiketli nesnelerin ne kadar doğru bir şekilde tahmin edildiğini ölçebilmesidir.
  - + Özellikle dengesiz veri kümelerinde önemli olan pozitif etiketli nesnelerin performansını ölçmek için kullanışlıdır.
  - Negatif etiketli nesnelerin performansını yansıtmaz.
  - Dengesiz veri kümelerinde yüksek recall değerleri, yanıltıcı olabilir.

$$Recall = \frac{TP}{TP + FN}$$

Recall, genellikle dengesiz veri kümelerinde kullanılır ve pozitif etiketli nesnelerin ne kadar doğru bir şekilde tahmin edildiğini ölçmek için kullanılır.

- Precision: Sınıflandırıcının pozitif olarak tahmin ettiği göstergelere sahip gerçek pozitiflerin oranıdır. Precision, sınıflandırıcının pozitif sınıfı yanlış olarak tahmin etme oranını ölçer.
  - + Özellikle dengesiz veri kümelerinde önemli olan pozitif etiketli nesnelerin performansını ölçmek için kullanışlıdır.

$$Precision = \frac{TP}{TP + FP}$$

Precision, genellikle dengesiz veri kümelerinde kullanılır ve doğru tahmin edilen pozitif etiketli nesnelerin ne kadar gerçekten pozitif etiketli olduğunu ölçmek için kullanılır.

- F1-score: F1-skoru (F1-score): Precision ve recall değerlerinin harmonik ortalamasıdır.
  - + Özellikle dengesiz veri kümelerinde önemli olan pozitif etiketli nesnelerin performansını ölçmek için kullanışlıdır.
  - Negatif etiketli nesnelerin performansını yansıtmaz.

- Precision ve recall arasındaki dengeyi ayarlamak için ekstra bir parametre gerektirir.

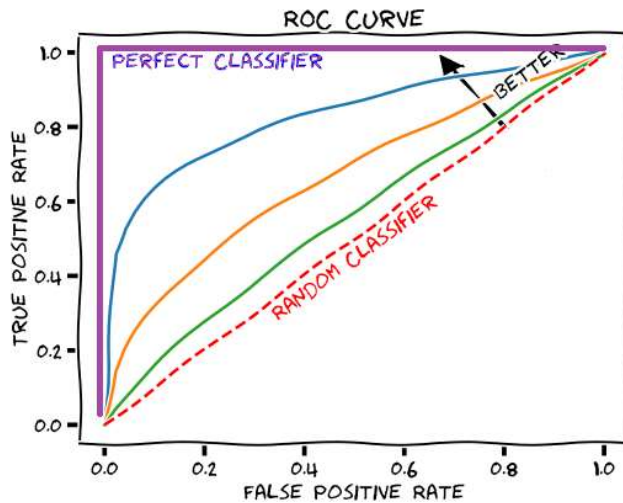
$$F - score = 2 * \frac{Precision * Recall}{Precision + Recall}$$

F1 score, genellikle dengesiz veri kümelerinde kullanılır ve modelin performansını hem doğru tahmin edilen pozitif etiketli nesnelerin ne kadar gerçekten pozitif etiketli olduğunu ölçtüğünü, hem de pozitif etiketli nesnelerin ne kadar doğru bir şekilde tahmin edildiğini ölçtüğünü aynı anda değerlendirmek için kullanılır.

- AUC-ROC (Receiver Operating Characteristic- Area Under the Curve): AUC-ROC, sınıflandırıcının pozitif sınıfı tahmin etme performansını ölçer. AUC-ROC, 0 ile 1 arasında bir değer alır ve bu değer ne kadar yüksekse, sınıflandırıcının performansı o kadar iyi demektir. Bu ölçüt, sınıflandırıcının pozitif ve negatif sınıflar arasındaki önem dengesinin değişmediği durumlarda kullanılabilir.

$$TPR = \frac{TP}{TP + FN}$$

$$FPR = \frac{FP}{FP + TN}$$



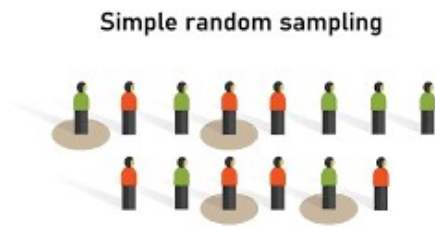
Şekil 2: AUC-ROC eğrisi



### 3 EĞİTİM VE TEST VERİLERİNİN SEÇİLMESİ

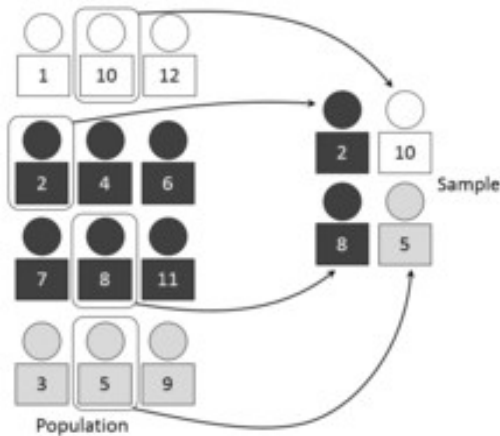
Bir sınıflandırıcının performansı, eğitim ve test setlerinin seçilme şekline göre etkilenebilir. Rastgele örnekleme, katmanlı örnekleme ve çapraz doğrulama gibi eğitim ve test setlerini seçmek için çeşitli yöntemler vardır. Her yöntemin kendine göre avantajları ve dezavantajları vardır ve yöntem seçimi sınıflandırıcının performansı üzerinde önemli bir etkiye sahip olabilir.

Rastgele örnekleme, örneklerin eğitim ve test kümeleri olarak kullanılmak üzere veri kümesinden rastgele seçildiği basit bir yöntemdir. Bu yöntemin ana avantajı basitliği ve uygulama kolaylığıdır. Ancak, örneklerin seçimi rastgele olduğundan ve tüm veri kümesini temsil etmeyebileceğinden, sınıflandırıcının performansında yüksek bir varyansa yol açabilir.



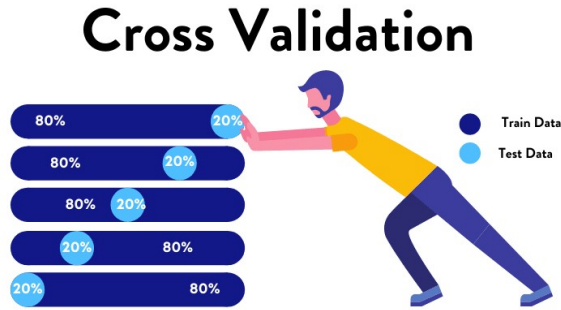
Şekil 3: Rastgele örnekleme örneği

Katmanlı örnekleme, eğitim ve test kümelerinin sınıf dağılımının tüm veri kümesinin sınıf dağılımını temsil etmesini sağlamak için örneklerin seçildiği bir yöntemdir. Bu yöntem, eğitim ve test setlerinin tüm veri setini temsil etmesini sağladığından, sınıflandırıcının performansındaki varyansı azaltmaya yardımcı olabilir. Ancak, özellikle veri seti dengesiz olduğunda mükemmel bir sınıf dağılımı elde etmek her zaman mümkün olmayabilir.



Şekil 4: Katmanlı örnekleme örneği

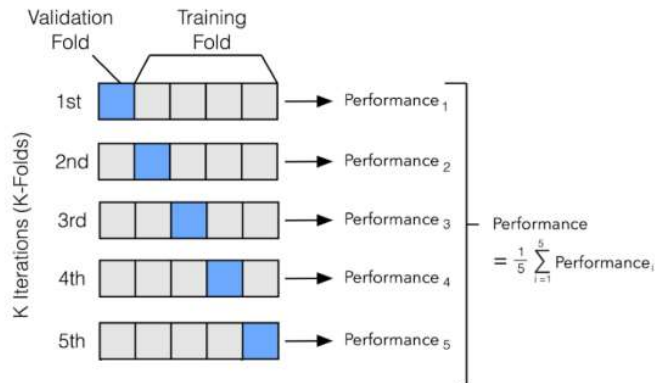
Cross-validation, veri kümesinin birden çok alt kümeye bölündüğü ve her bir alt kümenin bir test kümesi olarak, geri kalan alt kümelerin ise eğitim kümesi olarak kullanıldığı bir yöntemdir. Bu yöntem, belirli eğitim ve test kümeleri seçiminden etkilenme olasılığı daha düşük olduğundan, sınıflandırıcının performansının daha sağlam bir tahminini sağlayabilir. Ancak, büyük veri kümeleri için hesaplama açısından maliyetli olabilir.



Şekil 5: Cross validation örneği

K-fold cross validation, bir makine öğrenimi modelinin performansını değerlendirme yöntemidir. Temel fikir, veri setini k eşit boyutlu "kıvrımlara" bölmek ve ardından her seferinde test seti olarak farklı bir katlama ve eğitim seti olarak kalan k-1 katlama kullanarak modeli k kez eğitmektir. Daha sonra modelin performansı, k iterasyon üzerinden performans ölçümlerinin ortalaması alınarak tahmin edilir.

K-fold cross validation'ın ana avantajı, modelin performansının daha sağlam bir tahminini sağlayabilen verilerin farklı alt kümelerinde eğitilmesine ve değerlendirilmesine izin vermesidir. Aynı zamanda birden fazla iterasyonda sonuçların ortalamasını alarak performans tahminlerindeki varyansın azaltılmasına yardımcı olabilir [6].



Şekil 5: K-fold Cross-validation örneği

#### 4 SONUÇ

Sınıflandırıcı performansını ölçmek için birçok metrik kullanılabilir. En yaygın olarak kullanılan metrikler arasında accuracy, recall, precision, F1 skoru ve ROC eğrisi bulunmaktadır. Accuracy, sınıflandırıcının gerçekten ne kadar doğru tahmin yaptığını ölçen bir metriktir. Recall, sınıflandırıcının gerçek pozitif nesnelerin kaç tanesini doğru olarak sınıflandırdığını ölçen bir metriktir. Precision, sınıflandırıcının pozitif olarak sınıflandırdığı nesnelerin gerçekte pozitif olanlarının oranını ölçen bir metriktir. F1 score, precision ve recall arasındaki dengeyi ölçen bir metriktir. Receiving Operating Characteristic (ROC) eğrisi, sınıflandırıcının performansının True positive rate (TPR) ve FALSE positive rate (FPR) arasındaki ilişkisini gösteren bir grafikdir.

Metrik seçimi uygulamanın özelliklerine ve veri setinin niteliklerine bağlıdır. Eğer sınıf dağılımı dengeli ise doğruluk iyi bir metriktir ancak dengesiz veri setlerinde duyarlılık, özgüllük ve F1 skoru daha bilgilidir.

Bir sınıflandırıcının performansı, eğitim ve test setlerinin seçilme şekline etkilenebilir. Rastgele örnekleme, katmanlı örnekleme ve çapraz doğrulama gibi eğitim ve test setlerini seçmek için çeşitli yöntemler vardır. Her yöntemin kendine göre avantajları ve dezavantajları vardır ve yöntem seçimi sınıflandırıcının performansı üzerinde önemli bir etkiye sahip olabilir.

## KAYNAKÇA

1. <https://bilgisayarkavramlari.com/2013/03/31/siniflandirma-classification/> (Erişim: 09.01.2023)
2. <https://towardsdatascience.com/understanding-confusion-matrix-a9ad42dcfd62>
3. <https://medium.com/deep-learning-turkiye/s%C4%B1n%C4%B1fland%C4%B1rma-problemlerindeki-metrikler-33ee5f30f8eb>
4. [https://medium.com/@gulcanogundur/do%C4%9Fruluk-accuracy-kesinlik-precision-duyarl%C4%B1l%C4%B1k-recall-ya-da-f1-score-300c925feb38#:~:text=Duyarl%C4%B1l%C4%B1k%20\(Recall\)%20ise%20Positive%20olarak,bize%20yard%C4%B1mc%C4%B1%20olacak%20bir%20metriktir.](https://medium.com/@gulcanogundur/do%C4%9Fruluk-accuracy-kesinlik-precision-duyarl%C4%B1l%C4%B1k-recall-ya-da-f1-score-300c925feb38#:~:text=Duyarl%C4%B1l%C4%B1k%20(Recall)%20ise%20Positive%20olarak,bize%20yard%C4%B1mc%C4%B1%20olacak%20bir%20metriktir.)
5. <https://bilgisayarkavramlari.com/2010/09/30/f1-degerlendirme-f1-scoring/>
6. <https://medium.com/kodluyoruz/makine-ogrenmesi-modelleri-icin-veri-bolme-islemi-3b517ed74e37>