

INTRODUCTION TO ARTIFICIAL INTELLIGENCE DERSİ

ARAŞTIRMA ÖDEVİ

FATİH TALHA TÜMER – 191180081

İÇİNDEKİLER

[İÇİNDEKİLER i](#_Toc124169003)

[ŞEKİLLER LİSTESİ ii](#_Toc124169004)

[1. GİRİŞ 1](#_Toc124169005)

[2 SINIFLANDIRICILARI DEĞERLENDİRİLMESİNDE KULLANILAN ÖLÇÜTLER 2](#_Toc124169006)

[2.1. Ölçütler 2](#_Toc124169009)

[3 EĞİTİM TEST 3](#_Toc124169010)

[4 SONUÇ 4](#_Toc124169011)

[KAYNAKÇA 5](#_Toc124169012)

# ŞEKİLLER LİSTESİ

[Şekil 1: Karmaşıklık matrisi 2](#_Toc124172932)

[Şekil 2: AUC-ROC eğrisi 4](#_Toc124172933)

# GİRİŞ

# SINIFLANDIRICILARIN DEĞERLENDİRİLMESİNDE KULLANILAN ÖLÇÜTLER

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 “Desicion 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].

Table

Description automatically generated

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 nagative (Tip 2 hata): Negatif tahmin edildi ve sonuç yanlış
* TP + FP = Modelin pozifif 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.
* Error rate: Tüm tahminlerdeki hata oranını ölçmeye yarar. Genel performansı ölçmek için kullanılır ancak modelin performansını değerlendirmek için tek başına yeterli değildir.
* Recall: Sınıflandırıcının pozitif olarak tahmin ettiği göstergelere sahip gerçek pozitiflerin oranıdır. Recall sınıflandırıcının pozitif sınıfı doğru olarak tahmin etme becerisini ölçer.
* Precision: Sınıflandırıcının pozitif olarak tahmin ettiği göstergelere sahip gerçek pozitiflerin oranıdır. Precision, sınıflandırıcının pozitif sınıfı yanlış olarak tahmin etme oranını ölçer
* F1-sore: F1-skoru (F1-score): Precision ve recall değerlerinin harmonik ortalamasıdır.
* 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.

Chart, diagram

Description automatically generated

Şekil 2: AUC-ROC eğrisi

# EĞİTİM VE TEST VERİLERİNİN SEÇİLMESİ

* Random sampling: This involves randomly selecting a sample of data from the dataset to use as the training set, and using the remaining data as the test set. This is a simple and easy-to-implement approach, but it may not always lead to a representative sample of the data.
* Stratified sampling: This involves sampling the data in such a way that the proportions of different classes in the sample are the same as in the overall dataset. This can be useful when the classes are imbalanced (i.e., one class is much more common than the other).
* K-fold cross-validation: This involves dividing the data into a number of "folds," and then using one fold as the test set and the rest as the training set. This process is repeated K times, with a different fold serving as the test set each time. The final model is trained on the entire dataset and the performance is evaluated on the average performance across all K folds.
* Leave-one-out cross-validation: This is a special case of K-fold cross-validation, where K is equal to the number of samples in the dataset. This approach can be computationally expensive, but it can be useful for small datasets.
* Time series data: For time series data (i.e., data that is collected over time), it is important to ensure that the training and test sets are contiguous and non-overlapping. One approach is to use a "rolling window" to split the data into training and test sets. This involves selecting a window of consecutive time steps as the training set, and using the data immediately following the window as the test set. The window is then "rolled" forward by a certain number of time steps and the process is repeated.

# SONUÇ

# 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>.
5. <https://bilgisayarkavramlari.com/2010/09/30/f1-degerlendirme-f1-scoring/>