

Karmaşıklık Matrisi (Confusion Matrix)

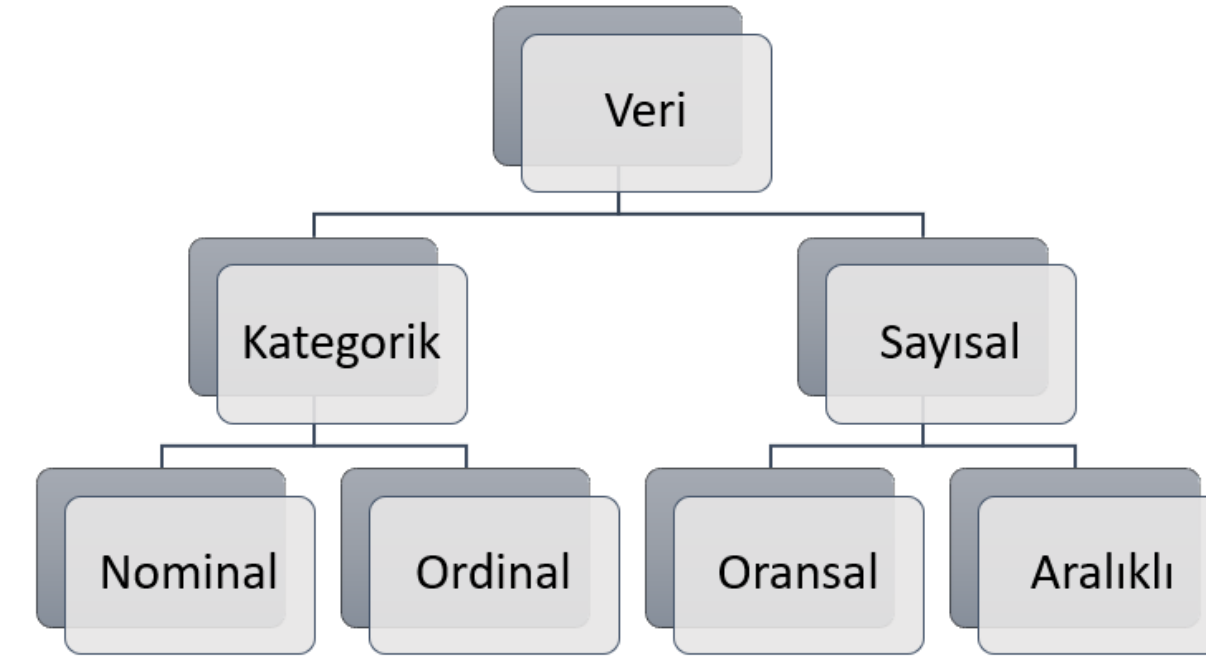
Karmaşıklık Matris(Confusion Matrix)

- Kurulan bir sınıflandırma modelinin sonuçlarını yorumlayabilmek, gerçek ve tahmin edilen değerler arasında yer alan ilişkideki hataları çapraz olarak inceleyebilmek için Confusion Matrix kullanılır.

***tipik olarak denetimli bir öğrenme algoritması olan bir algoritmanın performansının görselleştirilmesine izin veren özel bir tablo düzenidir

Sınıflandırma Problemleri İçin Karmaşıklık Matrisi

- Kategorik veriler sınıflanarak ifade edilen verilerdir, nitelik belirtirler. Ülke, cinsiyet, renk kategorik veriyken yaş, hız, sıcaklık numerik veridir. Sınıflandırma yaparken kategorik veri elde etmeye çalışırız.
- Sınıflandırma (classification) algoritmaları ikili (binary) ya da ikiden fazla (multiclass) kategorik değerleri sınıflandırmayı amaçlamaktadır. Bu amaç doğrultusunda ortaya çıkan hataların değerlendirilmesi ve doğruluk oranlarının en efektif biçimde ortaya konulabilmesi için Karmaşıklık Matrisleri(Confussion Matrix) kullanılıyor.



Yaş	Kilo	Boy	Cinsiyet
22	56	160	K
22	95	189	E
23	85	190	E
24	65	170	K
21	93	186	E
20	45	156	K

Confusion Matrix Gösterimi

Örnek: Kanser Teşhisi

Tahmin

Gerçek

n	Tahmin C1	Tahmin C2	Toplam
Gerçek C1	TP	FN	
Gerçek C2	FP	TN	
Toplam			

- True Positive(TP)
- True Negatives(TN)
- False Positive(FP)
- False Negatives(FN)

- True Positive ve True Negative modelin doğru olarak tahminlediği, False Positive ve False Negative ise modelin yanlış olarak tahminlediği alanlardır.
- TP, TN : Diagonal Elemanlar.



Confusion Matrix

Accuracy, Precision, Recall

Accuracy(Doğruluk)

Tahmin

Gerçek

n	Gerçek C1	Gerçek C2	Toplam
Tahmin C1	TP	FN	
Tahmin C2	FP	TN	
Toplam			

- Doğru tahminde bulunduğumuz alanları toplam veri kümesine(n) bölümü ile elde edilir
- $A = (TP + TN) / (TP + TN + FP + FN)$
- Bazı durumlarda modelin başarımını test etmek için yeterli değildir. Bu nedenler diğer metriklere de ihtiyaç duyulur.

Precision(Kesinlik)

Tahmin

Gerçek

n	Gerçek C1	Gerçek C2	Toplam
Tahmin C1	TP	FN	
Tahmin C2	FP	TN	
Toplam			

- Pozitif tahmin değerinin toplam pozitif değerler bölünmesi ile elde ediliyor

- $P = TP / (TP + FP)$

Recall(Duyarlılık)

Tahmin

Gerçek

n	Gerçek C1	Gerçek C2	Toplam
Tahmin C1	TP	FN	
Tahmin C2	FP	TN	
Toplam			

- Positive olarak tahmin etmemiz gereken işlemlerin ne kadarını Positive olarak tahmin ettiğimizi gösteren bir metriktir.

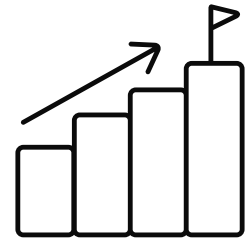
- $R = TP / (TP + FN)$

F1-Skor(F1-Score)

Gerçek	Tahmin		
	n	Gerçek C1	Gerçek C2
	Tahmin C1	TP	FN
	Tahmin C2	FP	TN
Toplam			

F-skoru: modelin Precision ile Recall birleştirmenin bir yoludur. Bu iki metriğin harmonik ortalaması olarak tanımlanır.

$$F_1 = \frac{2}{\frac{1}{\text{recall}} + \frac{1}{\text{precision}}} = 2 \times \frac{\text{precision} \times \text{recall}}{\text{precision} + \text{recall}}$$
$$= \frac{2 \times \text{tp}}{\text{tp} + \text{fp} + \text{fn}}$$



- <https://yigitsener.medium.com/>

- https://scikit-learn.org/stable/auto_examples/model_selection/plot_confusion_matrix.html#sphx-glr-auto-examples-model-selection-plot-confusion-matrix-py



- https://scikit-learn.org/stable/auto_examples/model_selection/plot_confusion_matrix.html#sphx-glr-auto-examples-model-selection-plot-confusion-matrix-py

- <https://ai.plainenglish.io/understanding-confusion-matrix-and-applying-it-on-knn-classifier-on-iris-dataset-b57f85d05cd8>



Referanslar