

Denetimli Öğrenmede Metriklerin Önemi

Makine öğrenimi genel olarak denetimli (supervised), denetimsiz (unsupervised) ve pekiştirmeli (reinforcement) öğrenme olarak 3 sınıfa ayrılır. Bu yazıda ve repoda supervised learning üzerine yoğunlaşıyor olacağız. Makine öğrenmesinde modellerin başarısını değerlendirmek için çeşitli metrikler kullanılır. Supervised learning algoritmalarına başlamadan önce, sonuçları nasıl yorum yapacağımızı öğrenmemiz gerekiyor, ki daha mantıklı yorumlar yapıp nerede ne kullanacağımızı daha net bir biçimde anlayıp görebilelim. Bu yazıda denetimli öğrenmede metrikleri öğrenip, önemini tartışıyor olacağız.

Supervised Learning

Öncelikle supervised learning ile başlayalım. Supervised learning, adından da belli olduğu üzere, makine öğrenmesinin denetimli öğrenme olarak bilinen kısmına denir. Peki neden "denetimli" öğrenme?

Denetimli öğrenme, modelin önceden etiketlenmiş hedef değişkenlerle eğitilmesi ve test edilmesi sürecine verilen addır. Model etiketli verileri kullanarak kalıplar ve bağlantılar oluşturarak öğrenir. Bu öğrenmeyi daha önce hiç görmediği ve etiketsiz olan verilerle test ederek nasıl ve ne kadar öğrendiğini ölçeriz. Supervised learning en genel tanımla kabaca ikiye bölünür:

1. Regression (Regresyon): Devamlı sayısal (continuous) değerleri olan hedef değişkenleri tahmin etmede kullanılır, input olarak daha çok sayısal değerler alır, bu nedenle sınıfları da encode ederek kullanırız.

Örneğin: [House Prices Advanced Regression Techniques](#)

2. Classification (Sınıflandırma): Kategorik (categorical) hedef değişkenleri tahmin etmek için kullanılır. Model, verilen girdilere dayanarak verileri belirli sınıflara ayırır. Girdi değişkenleri sayısal veya kategorik olabilir, kategorik değişkenler genellikle one-hot encoding veya label encoding ile işlenir.

Örneğin: [Titanic – Machine Learning from Disaster](#) 

Bu yazıda, aşağıdaki başlıklarda yer alan ilgili metrikleri, nasıl hesaplandıklarını, ne anlama geldiklerini ve hangi durumlarda hangi metriklerin daha uygun olduğunu analiz edeceğiz.

Temel Metrikler

Regresyon:

- Ortalama Mutlak Hata (Mean Absolute Error - MAE)
- Ortalama Kare Hata (Mean Squared Error - MSE)
- Kök Ortalama Kare Hata (Root Mean Squared Error - RMSE)
- R-kare (R^2)

Sınıflandırma:

- Doğruluk (Accuracy)
- Hassasiyet (Precision)

- Duyarlılık (Recall) ve F1 Skoru
- Kesinlik-Hatırlama Eğrisi ve ROC Eğrisi (AUC-ROC)
- Hata Matrisi (Confusion Matrix)

Hadi incelemeye başlayalım!

Regresyon Metrikleri

1. Mean Absolute Error – MAE (Ortalama Mutlak Hata):

MAE (Mean Absolute Error), bir modelin tahminleri ile gerçek değerler arasındaki farkların mutlak değerlerinin ortalamasını alarak hata büyüklüğünü ölçen bir metriktir. L1 kayıp fonksiyonu olarak da bilinir ve genellikle regresyon problemlerinde kullanılır. MAE, hataların büyüklüğünü doğrudan yorumlamayı kolaylaştırarak model performansını değerlendirmeye yardımcı olur ([C3 AI, 2024](#)).

$$MAE = \frac{1}{N} \sum_{i=1}^N |y_i - \hat{y}_i|$$

Formülü kullanılarak hesaplamaları yapılır.

2. Mean Squared Error – MSE (Ortalama Kare Hata):

MSE (Mean Squared Error), bir modelin tahminleri ile gerçek değerler arasındaki farkların karesinin ortalamasını alarak hesaplanan bir hata metriğidir. Hataları kare aldığı için büyük hatalara daha fazla ağırlık verir, bu yüzden aykırı değerlere karşı hassastır. Ayrıca, diferansiyellenebilir bir metrik olduğu için optimizasyon algoritmalarında yaygın olarak kullanılır. Ancak, hataların karesi alındığı için orijinal birimlerden farklı bir ölçekte ifade edilir, bu da yorumlamayı zorlaştırabilir ([NVIDIA Technical Blog, 2023](#)).

$$MSE = \frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2$$

3. Root Mean Squared Error – RMSE (Kök Ortalama Kare Hata)

RMSE (Root Mean Squared Error), MSE'nin karekökü alınarak hesaplanan bir hata metriğidir. Bu işlem, hatayı hedef değişkenle aynı ölçeğe getirerek yorumlamayı kolaylaştırır. Ancak, RMSE'nin belirli bir değeri doğrudan ortalama hata miktarını ifade etmez. MSE ile benzer özelliklere sahip olup, model optimizasyonunda aynı sonuçları verir ([NVIDIA Technical Blog, 2023](#)).

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2}$$

4. R-squared - R² (R kare)

R-kare (R²), bir modelin bağımlı değişkendeki (hedef değişken/label) varyansın ne kadarını açıkladığını gösteren bir metriktir. RSS (kalıntı kareler toplamı) ve TSS (toplam kareler toplamı) kullanılarak hesaplanır. R²'nin yüksek olması modelin veriye daha iyi uyum sağladığını gösterse de tek başına yeterli bir değerlendirme ölçütü değildir. Özellikle fazla değişken eklemek R²'yi artırabilir, ancak bu her zaman modelin daha iyi olduğunu göstermez. Bu durumu düzeltmek için düzeltilmiş R² kullanılır ([NVIDIA Technical Blog, 2023](#)).

$$R^2 = 1 - \frac{RSS}{TSS} = 1 - \frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{\sum_{i=1}^N (y_i - \bar{y})^2}$$

5. Diğer Regresyon Metrikleri

- 5.1. **Mean Absolute Percentage Error (MAPE):** MAPE tahmin hatalarını yüzdelik olarak ifade eden ve farklı ölçekteki verilerle karşılaştırılabilen popüler bir metrik olup, gerçek değerlerin sıfır olduğu durumlarda belirsizleşir. Ayrıca, negatif hataları (gerçek değerlerden yüksek tahminler) pozitif hatalardan (düşük tahminler) daha ağır cezalandırarak asimetrik bir yapıya sahiptir.

$$MAPE = \frac{1}{N} \sum_{i=1}^N \left| \frac{y_i - \hat{y}_i}{y_i} \right|$$

- 5.2. **Symmetric Mean Absolute Percentage Error (SMAPE):** SMAPE, MAPE'nin asimetrik yapısını düzeltmeye çalışan bir alternatif olup, 0% ile 200% arasında sınırlıdır ancak sıfıra yakın değerlerde stabil olmayan sonuçlar verebilir. SMAPE, aynı zamanda negatif hataları daha fazla cezalandırarak, genellikle düşük tahminlerin daha büyük cezalar almasına yol açar.

$$sMAPE = \frac{1}{N} \sum_{i=1}^N \frac{|y_i - \hat{y}_i|}{(y_i + \hat{y}_i)/2}$$

- 5.3. **Mean Squared Log Error (MSLE)**

- 5.4. **Root Mean Squared Log Error (RMSLE)**

- 5.5. **Akaike Information Criterion (AIC)**

Peki bu sık kullanılan regresyon metrikleri daha çok hangi alanlarda kullanılır?

- MAE (Mean Absolute Error): Hataların eşit şekilde değerlendirilmesi ve **uç değerlerin etkisinin minimize edilmesi** istendiğinde tercih edilir.
- MSE (Mean Squared Error): **Büyük hataların daha fazla cezalandırılması gerektiğinde** ve **modelin hassasiyetinin ön planda** olduğu durumlarda kullanılır. Ama, bu insan okuması ve yorumlamasına çok açık olmadığı için daha büyük çoğunlukla bunun karekökünü alarak (RMSE) okuma ve değerlendirme yaparız.
- RMSE (Root Mean Squared Error): **Hataların birimlerini anlamlı kılmak** ve daha kolay yorumlanabilir bir sonuç elde etmek istendiğinde tercih edilir.
- R² (Coefficient of Determination): **Modelin bağımsız değişkenlerle bağımlı değişkeni ne kadar iyi açıkladığı** görmek istendiğinde kullanılır.
- MAPE (Mean Absolute Percentage Error): **Tahminlerin yüzdelik hata oranlarını** göstererek, farklı veri setlerini karşılaştırmak için kullanılır.
- SMAPE (Symmetric Mean Absolute Percentage Error): Asimetrik hataların etkisini azaltmak ve aşırı tahmin hatalarını dengelemek gerektiğinde tercih edilir.

Sınıflandırma Metrikleri

Sınıflandırma metriklerini incelemeden önce Karışıklık Matrisini (Confusion Matrix) inceleyelim.

Karışıklık matrisi, sınıflandırma algoritmalarının performansını değerlendirmek için kullanılan temel bir araçtır. Bu matris, modelin doğru ve yanlış sınıflandırmalarını sayısal olarak gösterir. Aynı zamanda, her bir sınıf için yapılan tahminlerin özetini sunar ve modelin hatalarını daha ayrıntılı bir şekilde analiz etmeyi sağlar.

Bir **ikili sınıflandırma** problemi için karışıklık matrisi aşağıdaki gibi tanımlanır:

1. Confusion Matrix (Karışıklık Matrisi)

		Predicted Values	
		Positive	Negative
Actual Values	Positive	TP	FN
	Negative	FP	TN

İkili sınıflandırma problemi karışıklık matrisi

Matrisin bileşenlerini basitçe beraber inceleyelim. Başlangıçta kolay gibi görünen, çalıştıkça karmaşıklaştıran ama örnekleri inceledikçe kolaylaşan bu matrisin bileşenleri aslında neleri ifade ediyor? Burada aradığımız sınıftan pozitif sınıf olarak bahsedeceğiz (mesela bir hastalık veri setinde hastanın hasta olması, Titanic veri setinde survived = 1 olması gibi).

1. **True Positive (TP):** Modelin pozitif sınıfı doğru tahmin etmesi.
2. **True Negative (TN):** Modelin negatif bir sınıfı doğru tahmin etmesi.
3. **False Positive (FP):** Modelin pozitif sınıfı yanlış bir şekilde tahmin etmesi.
4. **False Negative (FN):** Modelin negatif sınıfı yanlış bir şekilde tahmin etmesi.

Çok karışık görüldüğünün farkındayım, hadi birer örnek kullanıp bu karışıklık matrisini çözüm matrisine çevirelim. Örneğin, bir spam mail/mesaj tespit etme veri setinde aradığımız özellik (bağımlı değişken/label) e postanın **spam olması** olsun. Bu durumda spam olması bizim için pozitif sınıf olacaktır. Aşağıdaki örneği inceleyelim.

		Predicted	
		Spam	Non-spam
Actual	Spam	600 (True positive)	300 (False negative)
	Non-spam	100 (False positive)	9000 (True negative)

Burada unutmamamız gereken: **True** değerlerin **doğru tahmin**; **positive** değerlerin **doğru sınıf** belirttiğini aklımızdan çıkarmamalıyız. Yoksa karışıklığa yol açabilir.

1. **True Positive (TP):** 600 adet TP burada gerçekten **spam** olan bir mailin doğru (True) bir şekilde **spam** olarak model tarafından tahmin edildiğini belirtiyor.
2. **True Negative (TN):** 9000 adet TN burada doğru bir şekilde **spam olmayan (negatif)** maillerin tahmin edildiğini gösteriyor.
3. **False Positive (FP):** 100 adet FP, aslında **pozitif olan sınıfın (spam olan mail)**, yanlış olarak (**spam olmayan**) tahmin edildiğini gösteriyor. Yani 100 tane e posta aslında spam maili iken, modelin yanlış tahmini yüzünden spam değildir olarak etiketleniyor. Bu durum ise False Positive olarak adlandırılır.

Bunun olası bir tehlikesi şu şekilde örneklendirilebilir: Örneğin bir kanser, tümör veya benzeri tehlikeli derecede önem taşıyan tıbbi veri setlerinde False oranları arttıkça (FP ve FN) bu sağlık açısından ciddi tehlike ve sorumluluk arz etmektedir. Eğer bir makine öğrenmesi modeli kanser olan bir bireyi (pozitif sınıf) kanser değil (false tahmin - negatif sınıf) şekilde etiketlerse bu, hasta açısından ciddi bir sorun teşkil edecektir. Bu nedenle özellikle sağlık alanlarında geliştirilen modellerde birden fazla metrik detaylı incelenerek kesin kanıya varılmalıdır.

4. **False Negative (FN):** 300 adet FN, negatif olan bir sınıfın yanlış olarak tahmin edilmesini simgelediğinden, burada 300 aden spam olmayan e postanın spam olarak işaretlendiğini görmekteyiz.

Burada tıbbi veri setinden örneklendirmemiz gerekirse, False Negative aslında **kanser olmayan** birinin **kanser** olarak etiketlenmesi sonucu gereksiz panik ve telaş oluşturacaktır. Yani aslında bu veri setinde çok ciddi bir tehlike arz etmemektedir. Ama durum her veri seti için aynı olmayacağından dikkatli kullanmakta fayda var.

True-False ile Positive-Negative değerlerini karıştırmamalıyız. **True** bizim için her zaman iyidir ve **doğru tahmin edilen değerleri** verir. False ise her zaman yanlış tahminleri vereceğinden modelin ince ayar (fine-tuning) aşamasında diğer sınıflandırma metriklerini hesaplarken kullandığımız önemli detayları içerir. Yazının devamında bahsediyor olacağız.

Örnekler ile karışıklık matrisini daha net bir şekilde oturttuğumuza göre ne işe yaradığını ve nerelerde kullanabileceğimizi de inceleyelim.

1. **Model Performansını Değerlendirme:** Modelin doğruluğuna (accuracy) bakmak yerine, sadece doğru tahminlerin oranına değil, aynı zamanda yanlış tahminlerin türlerine de bakmamız gerektiğini gösterir. Örneğin, bir modelin sadece doğruluğu yüksek olabilir ama **yanlış pozitiflerin fazla olması**, yanlış alarm ([False Alarm Rate - Science Direct](#), Signal Processing, 2008) verme durumunu yaratabilir.
2. **Duyarlılık (Recall) ve Kesinlik (Precision) Hesaplama:** Karışıklık matrisi, bu iki metriği hesaplamak için temel bir kaynaktır. Duyarlılık, hastaları doğru tespit etme oranını, hassasiyet ise doğru şekilde hasta olarak sınıflandırılanların doğru olma oranını ölçer. Bu iki metrik, özellikle **sağlık gibi önemli uygulamalarda** çok kritik bir önem taşır.
3. **F1 Skoru Hesaplama:** F1 skoru, duyarlılık ve hassasiyetin **harmonik ortalamasıdır**. Karışıklık matrisi, F1 skorunu anlamak için gereken verileri sağlar. Yazının devamında detaylarından bahsediyor olacağız.

4. **Hata Türlerini Anlama:** Eğer modelinizin yanlış pozitif ya da yanlış negatif tahminleri çok fazla ise, bu hataların nedenini analiz edebilirsiniz. Bu, modelinizi geliştirmek için nerelere odaklanmamız gerektiğini gösterir.

Sonuç olarak, karışıklık matrisi sadece modelin genel başarısını görmekle kalmaz, aynı zamanda modelin hangi alanlarda zayıf olduğunu da anlamamıza yardımcı olur. Bu da modelin iyileştirilmesi için gerekli adımları atmamızı sağlar.

Karışıklık matrisini net bir şekilde anladığımıza göre şimdi diğer sınıflandırma metriklerini inceleyebiliriz.

1. **Doğruluk (Accuracy):** Doğruluk, sınıflandırma modellerinde **en yaygın kullanılan metrik** olup, doğru **tahminlerin toplam tahmin sayısına oranını** ifade eder. Bu metrik, **veriler dengeli olduğunda** (çoğunluk sınıfı ile azınlık sınıfı arasındaki fark çok büyük olmadığında) etkili bir şekilde kullanılabilir.

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

2. **Kesinlik (Precision):** Kesinlik, pozitif sınıf için doğru şekilde tahmin edilen örneklerin, toplam pozitif olarak tahmin edilen örneklere oranını ölçer. Bu metrik, **yanlış pozitiflerin maliyetinin yüksek olduğu durumlarda önemlidir** çünkü yalnızca doğru pozitif tahminlere odaklanır.

$$Precision = \frac{TP}{TP + FP} \quad \text{or} \quad \frac{\text{True Positive}}{\text{Predictive Results}}$$

3. **Duyarlılık (Recall) veya Hassasiyet (Sensitivity):** Duyarlılık, pozitif sınıfın doğru şekilde sınıflandırılan örneklerin oranını ölçer. **Veriler dengesiz olduğunda, özellikle azınlık sınıfının pozitif olduğu durumlarda** (örneğin Covid-19 veya diğer hastalık veri setlerinde) bu metrik tercih edilmelidir. **Çünkü yanlış bir pozitif hastayı negatif olarak tahmin etmek istenmez.**

$$Recall = \frac{TP}{TP + FN} \quad \text{or} \quad \frac{\text{True Positive}}{\text{Actual Results}}$$

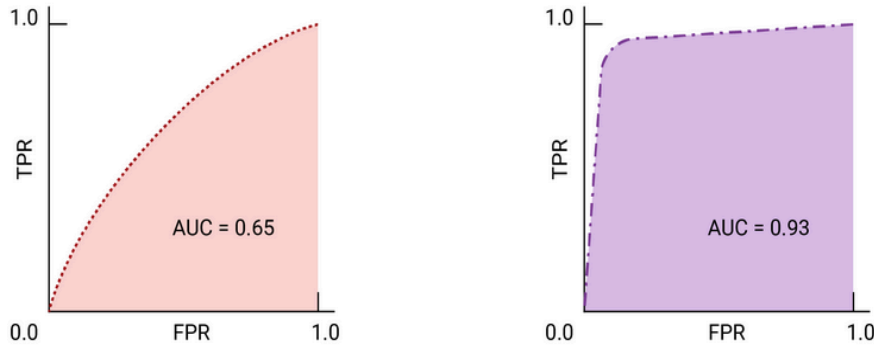
4. **Özgüllük (Specificity):** Özgüllük, recall'ın tersine, negatif sınıfın doğru şekilde sınıflandırılan örneklerinin oranını ölçer. **Veriler dengesiz olduğunda ve azınlık sınıfı negatif olduğunda** (örneğin, müşteri kaybı veri setlerinde) bu metrik kullanılması önerilir. Çünkü burada amaç, yanlış bir negatif örneği pozitif olarak sınıflandırmamak ve negatif sınıfı doğru şekilde tespit etmektir. Ama recall ölçeği, sensitivity'den daha sık kullanılır.

$$\text{Specificity} = \frac{TN}{TN + FP}$$

5. **F1 Skoru:** F1 skoru, duyarlılık ve hassasiyet değerlerinin harmonik ortalamasını temsil eder. Bu metrik, her iki ölçütün dengeli bir şekilde değerlendirilmesini sağlar ve genellikle hem hatalı pozitifleri hem de hatalı negatifleri dikkate alarak daha kapsamlı bir değerlendirme yapar.

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

6. **ROC-AUC Eğrisi (Curve):** ROC-AUC eğrisi, modelin tüm sınıflandırma eşiklerinde (threshold) doğruluk oranını ölçer ve modelin pozitif sınıfı ne kadar iyi ayırt edebildiğini gösterir. AUC değeri, 0 ile 1 arasında değişir ve 1'e yaklaştıkça modelin başarısı artar. **Özellikle dengesiz veri setlerinde**, AUC (Eğrinin altındaki alan) skoru modelin genel performansını daha dengeli bir şekilde değerlendirmek için tercih edilir (Buczak, S. et al., 2016).



AUC, iki farklı modelin performansını karşılaştırmak için kullanışlı bir ölçümdür, ancak veri seti yaklaşık olarak dengeli olduğu sürece. Eğri altındaki alanı (AUC değeri) daha büyük olan model genellikle daha iyi olanıdır. ([Google Developers - Machine Learning -Classification: ROC and AUC](#), n.d.).

Bu metriklerin araştırması sırasında kullanılan kaynaklar:

([Understanding Evaluation Metrics in Classification Modeling | Towards Data Science](#), 2021;
[What is A Confusion Matrix in Machine Learning? The Model Evaluation Tool Explained | DataCamp](#), 2024;
[Artificial General Intelligence - C3 AI](#), n.d.,
[Medium - Doğruluk \(Accuracy\) , Kesinlik\(Precision\) , Duyarlılık\(Recall\) ya da F1 Score ?](#), 2019).

Örnek Senaryolar Kullanarak Yorumlama

Örnek senaryolar kullanarak sadece 1 ölçeğe neden tek başına güvenmemeliyiz sorusunu beraber yanıtlayalım. Örneğin, kanser teşhisinde yüksek doğruluk oranı neden yanıltıcı olabilir? Öncelikle biraz düşünelim, sonra okumaya devam edelim.

Yeterince düşündüğümüze göre nedenini şimdi beraber ele alabiliriz. Kanser teşhisinde yüksek doğruluk oranı (accuracy), tek başına yanıltıcı olabilir çünkü accuracy, modelin tüm tahminleri içindeki doğru tahminlerin oranını gösterir **ancak sınıfların dağılımını ve modelin gerçek performansını tam olarak yansıtmaz**. Confusion matrix bileşenlerini kullanarak şu şekilde açıklayabiliriz:

- Hatalı negatifler (FN), kanser hastasına yanlışlıkla "sağlıklı" denirse, bu **ölümcül** sonuçlara yol açabilir.
- Hatalı pozitifler (FP), sağlıklı birine yanlışlıkla "kanser" teşhisi konulursa, **gereksiz tıbbi testler, stres ve mali yük** oluşturabilir.

Bu konuda hangi metriklerin kullanımı daha doğru olur?

Accuracy yerine, tıp alanında genellikle **recall, specificity, F1 skoru** ve **ROC-AUC (Receiver-operating characteristic curve - Area Under the Curve)** gibi metrikler tercih edilir.

- Recall, gerçekten hasta olanların kaçının doğru tespit edildiğini gösterir.
- Specificity, gerçekten sağlıklı olanların kaçının doğru tespit edildiğini gösterir. (Yani aslında recall ile aynı işi görür ama farklı bileşenleri kullanarak.)
- F1 Skoru, hem accuracy, hem precision hem de recall'ı dengeleyen bir metriktir.

Peki, Sınıflandırmada Precision mı yoksa Recall mı daha önemli?

Aslında bu soru, hangi tür hataların daha kritik olduğuna bağlı olarak değişir.

- Precision, modelin **tahmin ettiği pozitif sınıfların ne kadarının gerçekten doğru olduğunu** gösterirken, Recall ise **gerçek pozitiflerin ne kadarını doğru tespit edebildiğini** ölçer.
- **Precision önemli olduğunda, yanlış pozitiflerin maliyeti yüksektir.** Örneğin yukarıda da bahsettiğimiz üzere, spam e-posta tespiti gibi bir senaryoda, spam olmayan e-postaların yanlışlıkla spam olarak sınıflandırılması (yani yanlış pozitif) istenmez. Bu durumda, Precision daha öncelikli olur.
- Recall daha önemli olduğunda, **yanlış negatiflerin (false negatives) maliyeti büyüktür.** Örneğin, kanser teşhisi gibi bir sağlık senaryosunda, kanserli hastaların yanlışlıkla sağlıklı olarak sınıflandırılması ciddi sonuçlar doğurabilir. Bu durumda, Recall ön planda tutulur.
- Sonuç olarak, hangi metrik daha önemli olacağı, **uygulama ve hata türüne göre değişir. Eğer önemli olan doğru pozitiflerin tespit edilmesi ise Recall, yanlış pozitiflerin minimize edilmesi ise Precision daha kritik olabilir.**

E-posta Spam Tespiti Örneğinde, Accuracy ile F1-Score Farkı Nedir?

Yazının sınıflandırma kısmında e posta spam detection örneğini detaylıca inceleyip ardından sınıflandırma ölçeklerini ele almıştık. Peki e posta spam örneğine accuracy ve f1-score farkı ne olabilir?

Kısaca bahsetmemiz gerekirse, accuracy **tüm doğru tahminlerin oranı** iken, f1-score **precision ve recall'un dengelenmiş halidir**. Spam gibi dengesiz veri setlerinde f1 skoru kullanmak bize daha anlaşılır ve detaylı bir inceleme fırsatı sunar.

Mean Absolute Error mı yoksa Root Mean Squared Error mı daha önemli?

- RMSE, daha büyük hataları daha fazla cezalandırır. Bu, modelin büyük hatalar yapmasının daha istenmediği durumlarda tercih edilir.
- Örneğin, **ev fiyatı tahmini** gibi bir durumda, **büyük fiyat hatalarının ciddi finansal sonuçlara yol açabileceği** için **RMSE** daha uygun olabilir.
- MAE ise **her hata büyüklüğünü eşit şekilde ele alır**, bu da onu daha basit ve anlaşılır bir metrik yapar. **Her hatanın eşit derecede önemli olduğu ve hatalarda genellendirme yapmak gerektiği zaman kullanılır**.
Örneğin, bir ürünün talep tahmini gibi senaryolarda, hataların büyüklüğüne değil, genelde hata oranına odaklanmak istenebilir.
- Genel olarak, **RMSE büyük hataların daha fazla etkisi olduğu senaryolarda** daha önemli olurken, **MAE daha düz ve anlaşılabilir** bir hata ölçümü sunar.

Hata toleransının düşük olduğu finansal tahminlerde hangi metrik daha anlamlıdır?

Hata toleransının küçük olması bizim için çoğu konuda önem teşkil eder ve bu da False tahmin değerlerini olabildiğince minimize etmemiz gerektiği anlamına gelmektedir. Bu da kullandığımız metriklere dikkat etmemiz gerektiği anlamına gelir, özellikle Finans, sağlık ve savunma sanayii gibi alanlarda proje geliştiriyorsak. Çoğunlukla model değerlendirmesi yaparken bir tane ölçek kullanmak doğru değildir ve finansal tahminlerde kullanılabilecek birden fazla metrik bulunmaktadır. Bunlardan en önemli iki tanesi şu şekildedir:

- RMSE, büyük hata cezalandırması nedeniyle yüksek risk içeren finansal tahminlerde daha anlamlıdır.
- MAPE, oransal hata ölçümü sağlar, farklı ölçeklerde kıyaslama yapmak için iyidir.