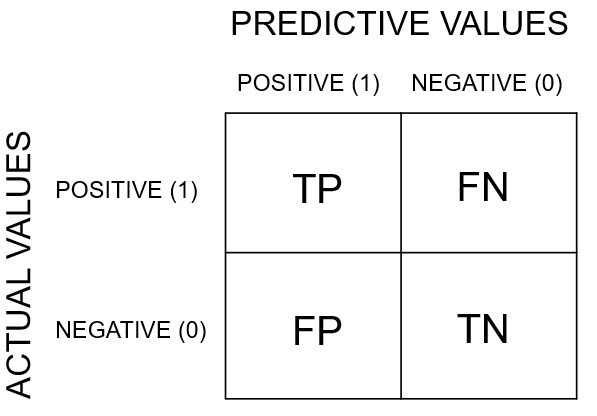
**SINIFLANDIRMA MODELLERİNDE BAŞARI KRİTERLERİ**

Düşünün ki 10.000 kişiden alınan çeşitli özniteliklere bakarak bu kişilerin kanser olup olmadığını tahmin etmek istiyorsunuz ve içlerinde 10 kişi gerçekten kanser. Biz herkese sağlıklı dersek. Accuracy (Doğruluk) = (9.990/10.000) \* 100= %99.9 olur. Rastgele bir tahminin doğruluk oranının bu kadar yüksek olması doğru yolda olduğumuzu gösterir mi?

**CONFUSION MATRIX (KARIŞIKLIK MATRİSİ)**

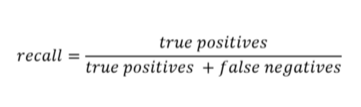
**Bu veri içerisinde;**

TP (True positive): Hastaya hasta demek.

FP (False positive): Hasta olmayana hasta demek.

TN (True negative): Hasta olmayana hasta değil demek.

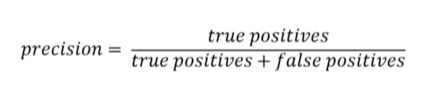
FN (False negative): Hasta olana hasta değil demek.

* **Recall (*Duyarlılık*): Hasta olanları doğru tespit etme oranı?**

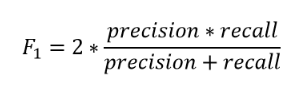
**Artısı:** Bazı anomali vakalarını doğru tespit etmek yanlış alarm üretmekten daha önemli. Diğer bir deyişle FN, FP’den daha kritik. Kanserli birini tespit edemeyip ölümüne neden olmaktansa kanser olmayan biri için yanlış tahmin yapıp onu hastaneye çağırmak daha kabul edilebilir.

**Eksisi:** Herkesi hastaneye çağırdık (Recall = 1) bütün kanserli vakaları bulduk ama alarmların çoğu yanlış (FP yüksek)

* **Precision (Kesinlik):**Hasta dediklerimizin gerçekten kaçı hasta?

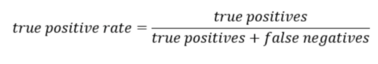
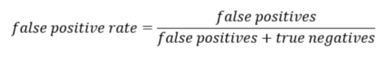
**Artısı:** Birine hasta demeden önce iyice düşünüp taşınmamızı sağlıyor. Herkes hasta dersek Precision = 10/10.000 = 0.001

**Eksisi:**Eğer bir kişiye hasta dedik ve o kişi gerçekten hasta (Precision = 1) ama kalan 9 kişiyi tespit etmedik. (FN yüksek)

Recall ve Precision iki önemli metrik ve aralarında trade-off var. Bununla baş edebilmek için F1 skoru kullanılıyor. **F1-skoru** ekstrem durumları cezalandırmak için aritmetik ortalama yerine **harmonik ortalamayı** kullanıyor. 10000 hasta olsun, bu hastalardan 10’unu kanser olma varsayalım. Böyle bir durumda Recall ve Precision kullanmadan yapacağımız tahminimizde herkese sağlıklı deriz ve Accuracy (doğruluk) oranımız **%99.9** olur ama tahminin Recall ve Precision oranları **0**.

* **ROC Curve (Receiver Operating Characteristic Curve)**

Sınıflandırma modellerinin başarı hesaplarında ROC Curve sık sık karşımıza çıkmaktadır. İlk başta karışık gelebilir temelde iki basit metriğe bakıyor.

* **TPR:** Kanser olan insanlara hangi olasılıkla doğru diye alarm veriyorum. (Recall aslında)
* **FPR:** Kanser olmayan insanlara hangi olasılıkla yanlış alarm veriyorum. (Bu metrik yeni daha önce görmedik.)

Bu iki metriği x ve y eksenlerine yerleştirerek çizginin altında kalan alanı hesaplıyoruz **(AUC — Area Under Curve). TPR: Y ekseni, FPR: X ekseni**

Her bir eğri bir modeli temsil ediyor. Eğri boyunca düşen threshold (sınır) değerlerine karşı TPR ve FPR oranları tespit ediliyor (o noktalar eğriyi oluşturuyor aslında.)

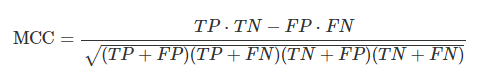
Her threshold için Precision, Recall değerlerini hesaplayıp F1 skorunu maksimum yapan thresholdu seçiyoruz.

Rastgele bir sınıflandırma (Random Classification) yapıldığında çizginin altında kalan alan 0.5 olmakta. Çizginin altında kalan alan ne kadar büyükse modelin başarı oranı o kadar yüksek demektir. Aslında bir modelin **F1 skoru** ne kadar yüksekse **çizginin altında kalan alan** da o kadar yüksek deyip işin içinden kolayca sıyrılabiliriz.

Şimdiye kadar bahsedilen problem FP ve FN durumlarına eşit yaklaşmasıdır. Halbuki kredi kartı dolandırıcılığını tespit edememek (FN) yanlış tespit yapıp müşteriyi aramaktan (FP) daha büyük maliyet içerir. Bununla başa çıkmak için çeşitli yöntemler var:

FN = k \* FP (örneğin k = 3). Böylece FN durumunun FP durumundan 3 kat daha problemli olduğunu modele ekleyebiliriz.

**MATTHEWS CORRELATION COEFFICIENT**

Matthews Korelasyon Katsayısı (MCC) -1 ile 1 arasında bir değere sahiptir; burada -1 tamamen yanlış bir ikili sınıflandırıcıyı gösterirken, 1 tamamen doğru bir ikili sınıflandırıcıyı gösterir. MCC'nin kullanılması, bir sınıflandırma modelinin / fonksiyonunun ne kadar iyi performans gösterdiğini ölçmenize olanak sağlar. ROC eğrisi ile benzer işlevi görür. Confusion Matrix kullanılarak elde edilen bir katsayıdır. 1’e ne kadar yakınsa bizim için o kadar iyidir.