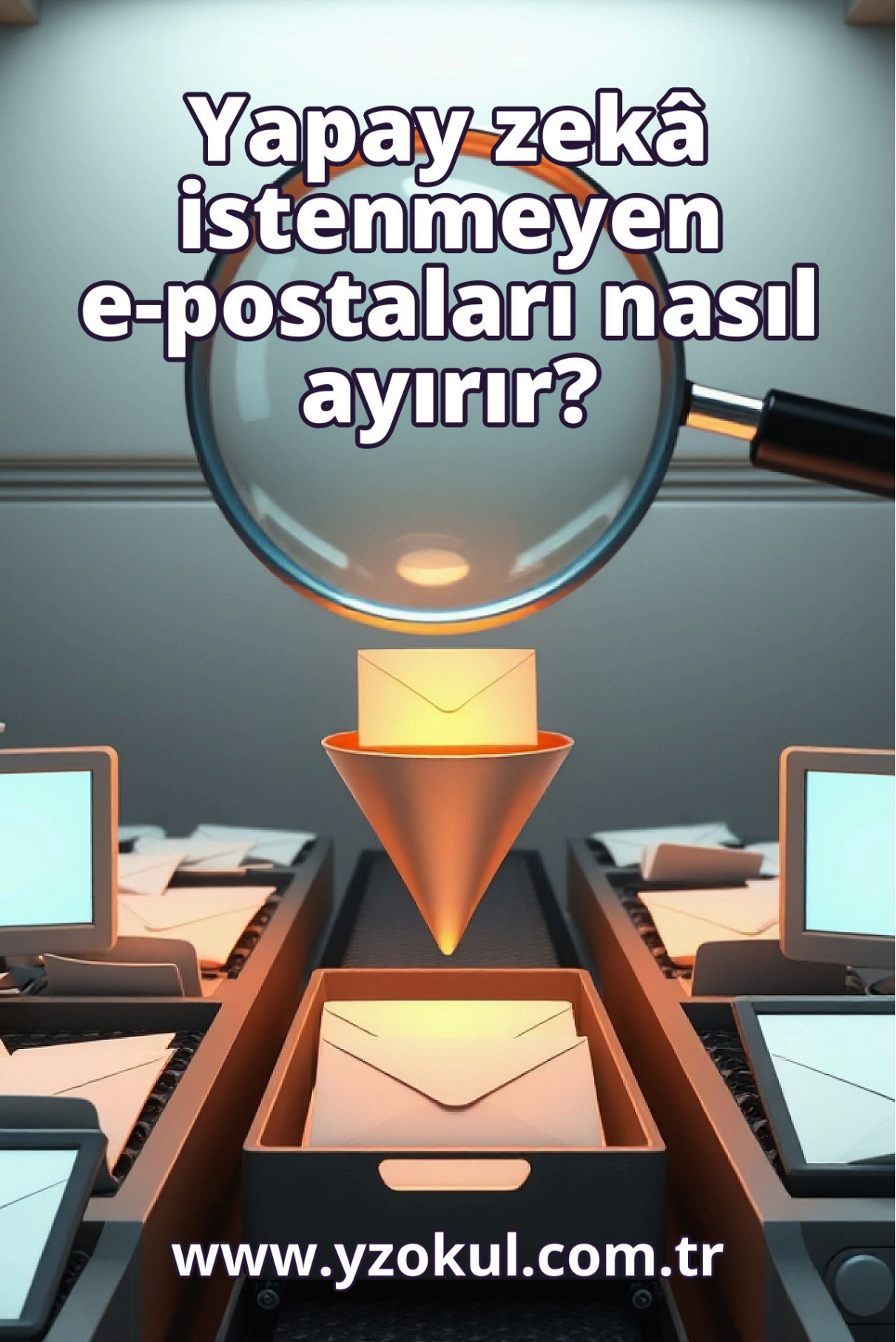

Yapay zekâ istenmeyen e-postaları nasıl ayırır?
www.yzokul.com.tr

Yapay zekâ istenmeyen e-postaları nasıl ayırır?



Yapay zekânın istenmeyen e-postaları nasıl ayırdığını merak edebilirsin. Kısa yanıt, bir iletideki birçok ipucuna baktığı ve geçmiş örneklere dayanarak iletinin istenmeyen olma olasılığını tahmin ettiğidir. Tek bir ipucuna göre karar vermez. Bunun yerine, karar vermeden önce birkaç sinyali birleştirir. Bu kitap her ipucunu ve bunların adım adım nasıl birlikte çalıştığını açıklar.



İstenmeyen e-posta, genellikle spam olarak adlandırılır ve istemediğin postadır. Çoğu zaman reklamdır ya da seni bir şeye tıklamaya veya bilgi paylaşmaya kandırma girişimidir. Bir e-posta filtresinin amacı bu iletileri okumak istediğin postadan ayırmaktır. İyi bir filtre gelen kutunu temiz tutarken normal iletilerin geçmesine izin verir. Spam'in ne olduğunu anlamak, filtrenin neden belirli ipuçlarını aradığını görmene yardımcı olur.



Bir filtre, postan için bir ayıklama kapısı gibi çalışır. Gelen her ileti bu kapıdan geçmeli ve ya gelen kutuna ya da spam klasörüne gönderilmelidir. Sistem, tam emin olmasa bile her bir ileti için karar vermek zorundadır. Karar, iletide bulunan ipuçlarına dayanır. Bu kitabın geri kalanı bu ipuçlarının neler olduğunu açıklar.



İlk ipucu ileti içeriğidir: e-postanın içindeki kelimeler ve ifadeler. Sistem metni okur ve istenmeyen postada normal postaya göre daha sık görünen kelimeleri ve kalıpları arar. Örneğin, hemen harekete geç veya sınırlı teklif gibi ifadeler spam'de sıkça görünebilir. Ancak sistem, daha sonra göreceğin gibi tek bir kelimeye güvenmez. Metnin genel kalıbına bakar.



İkinci ipucu gönderen bilgisidir. Sistem iletinin kim tarafından gönderildiğini, gönderen adresi ve bunun nasıl yazıldığı dahil olmak üzere kontrol eder. Garip veya bilinmeyen bir gönderen bir uyarı işareti olabilir. Örneğin, çok sayıda rastgele sayı içeren bir adres veya yanlış yazılmış bir şirket adı spam'i düşündürebilir. Sistem bu bilgiyi geçmiş iletilerden öğrendikleriyle karşılaştırır. Gönderen ayrıntıları her e-postanın daha tam bir resmini oluşturmaya yardımcı olur.



Üçüncü ipucu bağlantı özellikleridir. Sistem bir iletinin içindeki bağlantıları, özellikle gerçekte nereye gittiklerini inceler. Gerçek gibi görünen ama başka bir yere işaret eden bağlantıya aldatıcı bağlantı denir. Örneğin, bir bankaya gidiyor gibi görünen ama aslında farklı bir web sitesine giden bağlantı şüphelidir. Bağlantı özelliklerini kontrol etmek, seni tıklamaya kandırmaya çalışan iletileri yakalamaya yardımcı olur. Bağlantılar, diğer sinyallerle birleştirildiğinde güçlü bir ipucudur.



Sistem geçmiş örneklerden öğrenir. Ona, insanların zaten istenen veya istenmeyen olarak işaretlediği birçok eski ileti gösterilir. Bu örneklerden hangi ipuçlarının spam ile birlikte gitme eğiliminde olduğunu öğrenir. Örneğin, geçmişteki birçok spam iletisi aldatıcı bağlantılar içerdiyse, sistem bu ipucunu bir uyarı olarak ele almayı öğrenir. Örneklerden öğrenmek, sistemin her durum için açıkça programlanmadan bilgisini nasıl oluşturduğudur.



Sistem kesin olarak evet veya hayır demez. Bunun yerine bir olasılık tahmin eder; bu, iletinin istenmeyen olma olasılığını gösteren bir sayıdır. Örneğin, bir iletinin spam olma olasılığını yüzde 90 olarak tahmin edebilir. Bu sayı, sistemin bundan sonra ne yapacağına karar vermesine yardımcı olur. Olasılık, iletide bulunan tüm ipuçlarına dayanır.



İpuçları tek tek değil, birlikte tartılır. Tekrarlanan reklam kalıpları artı aldatıcı bir bağlantı güçlü bir sinyal oluşturabilir. Örneğin, birçok satış kelimesi ve garip bir web sitesine işaret eden bir bağlantı içeren bir ileti, bu ipuçlarından yalnızca birini içeren bir iletiden daha büyük olasılıkla spam'dir. İpuçlarını birleştirmek daha güvenilir bir tahmin verir. Sistem karar vermeden önce kanıtları toplar.



Tek bir kelime sonucu belirleyemez. Normal bir ileti şüpheli bir kelime içerebilir, bu yüzden sistemin daha fazla kanıta ihtiyacı vardır. Örneğin, bir arkadaşın bulduğu harika bir tekliften bahsedebilir ve yalnızca teklif kelimesi onu spam'e göndermemelidir. Sistem tek bir kelimeye değil, tüm iletiye bakar. Bu yüzden olasılığı tahmin etmeden önce birçok ipucunu birleştirir.



Sistem bir eşik, yani bir karar çizgisi kullanır. Belirli bir olasılığın üzerinde ileti spam klasörüne gider. Bu çizginin altında gelen kutuna gider. Bu çizginin nerede durduğu, filtrenin ne kadar katı olduğunu değiştirir. Düşük bir eşik, daha fazla iletinin spam olarak ele alınması anlamına gelirken yüksek bir eşik daha azının ele alınması anlamına gelir. Doğru eşiği seçmek bir dengedir.



Bazen normal bir ileti yanlışlıkla engellenir. Bu, ipuçları tesadüfen spam gibi göründüğünde olur. Örneğin, abone olduğun bir bülten birçok reklam kelimesi ve bir bağlantı içerebilir, bu yüzden filtre onu spam'e gönderebilir. Bu yaygın bir hatadır. İpuçları yanıltıcı olduğunda iyi eğitilmiş bir sistemin bile hata yapabileceğini gösterir.



Bazen istenmeyen bir ileti gelen kutuna ulaşır. Bu, ipuçları yeterince zararsız göründüğünde olur. Örneğin, bağlantı içermeyen sade bir dille yazılmış bir spam iletisi filtreden sıyrılabilir. Bu diğer yaygın hatadır. Sistemin her istenmeyen iletiyi yakalayamayacağını gösterir. Bazı spam'ler her zaman geçecektir.



İki hata zıt yönlere çeker. Filtreyi daha katı yapmak kaçan spam'i azaltır ama daha fazla normal postayı engeller. Daha gevşek yapmak engellenen normal postayı azaltır ama daha fazla spam'in geçmesine izin verir. Örneğin, eşiği düşürmek daha fazla spam yakalar ama en sevdiğin bülteni de yakalayabilir. Mükemmel bir ayar yoktur. Amaç, çoğu ileti için iyi işleyen bir denge bulmaktır.



İnsanlar filtrenin sonuçlarını gözden geçirerek gelişmesine yardımcı olur. Bir ileti yanlışlıkla engellendiğinde, onu spam değil olarak işaretleyebilirsin. Bu düzeltmeler sistemin öğrenebileceği yeni örnekler olur. Örneğin, engellenen bir bülteni gelen kutuna geri taşırsan sistem ipuçlarını ayarlayabilir. Zamanla bu incelemeler filtrenin daha iyi kararlar vermesine yardımcı olur. Öğrenme, sistem zaten kullanımdayken de devam eder.



Özetle, sistem birçok ipucunu birlikte değerlendirir ve geçmiş örneklere dayanarak bir olasılık tahmin eder. İleti içeriğine, gönderen bilgilerine ve bağlantı özelliklerine bakar. Sonucu tek bir kelime belirlemez. Bu nedenle bazı normal postalar engellenebilir ve bazı spam'ler geçebilir. Filtre bu hataları dengelemeyi ve geri bildirimle gelişmeyi amaçlar.