7.Evaluating Malware Detection Systems:Kötü Amaçlı Yazılım Tespit Sistemlerinin Değerlendirilmesi:

Önceki bölümde, makine öğreniminin kötü amaçlı yazılım algılayıcıları oluşturmanıza nasıl yardımcı olabileceğini öğrendiniz. Bu bölümde, kötü amaçlı yazılım tespit sistemlerinin nasıl çalışacağını tahmin etmek için gerekli temel kavramları öğreneceksiniz. Burada öğrendiğiniz fikirler, kurduğunuz herhangi bir kötü amaçlı yazılım tespit sistemini iyileştirmede çok önemli olacaktır, çünkü sisteminizin performansını ölçmenin bir yolu olmadan, onu nasıl iyileştireceğinizi bilemezsiniz. Lütfen bu bölümün temel değerlendirme kavramlarını tanıtmaya adanmış olmasına rağmen, Bölüm 8'in çapraz doğrulama gibi temel değerlendirme kavramlarını tanıtarak devam ettiğini unutmayın.

İlk olarak, algılama doğruluğu değerlendirmesinin arkasındaki temel fikirleri tanıtıyorum ve ardından performansını değerlendirirken sisteminizi yerleştirdiğiniz ortama ilişkin daha gelişmiş fikirleri ortaya koyuyorum. Bunu yapmak için, farazi bir kötü amaçlı yazılım tespit sisteminin bir değerlendirmesini size yönlendiriyorum.

->Four Possible Detection Outcomes: Dört Olası Tespit Sonucu

Bir yazılım ikili programında kötü amaçlı yazılım algılama sistemi çalıştırdığınızı ve ikili dosyanın kötü niyetli mi yoksa zararsız mı olduğu konusunda sistemin "görüşünü" aldığınızı varsayalım. Şekil 7-1'de gösterildiği gibi, dört olası sonuç ortaya çıkabilir.

Bu sonuçlar şu şekilde tanımlanabilir:

True positive: Gerçek pozitif İkili kötü amaçlı yazılımdır ve sistem kötü amaçlı yazılım olduğunu söyler.

False negative: Yanlış negatif İkili kötü amaçlı yazılımdır ve sistem kötü amaçlı yazılım olmadığını söylüyor.

False  positive: Yanlış pozitif İkili kötü amaçlı yazılım değil ve sistem kötü amaçlı yazılım olduğunu söylüyor.

True negative: Gerçek olumsuz İkili kötü amaçlı yazılım değildir ve sistem kötü amaçlı yazılım olmadığını söylüyor.

Gördüğünüz gibi, kötü amaçlı yazılım tespit sisteminizin yanlış sonuçlar üretebileceği iki senaryo vardır: yanlış negatifler ve yanlış pozitifler. Uygulamada, gerçek olumlu ve gerçek olumsuz sonuçlar arzuladığımız şeydir, ancak bunları elde etmek genellikle zordur.

Bu bölüm boyunca bu terimlerin kullanıldığını göreceksiniz. Aslında, tespit değerlendirme teorisinin çoğu bu basit kelime dağarcığı üzerine inşa edilmiştir.

->True and False Positive Rates:Doğru ve Yanlış Pozitif Oranlar

Şimdi, bir dizi zararlı yazılım ve kötü amaçlı yazılım kullanarak algılama sisteminin doğruluğunu test etmek istediğinizi varsayalım. Dedektörü her ikili üzerinde çalıştırabilir ve dedektörün tüm test seti boyunca size dört olası sonuçtan hangisini vereceğini sayabilirsiniz. Bu noktada, size sistemin doğruluğu hakkında genel bir fikir vermesi için bazı özet istatistiklere ihtiyacınız vardır (yani, sisteminizin yanlış pozitifler veya yanlış negatifler üretme olasılığı nedir).

Bu tür bir özet istatistik, test setinizdeki gerçek pozitiflerin sayısını test setinizdeki toplam kötü amaçlı yazılım örneği sayısına bölerek hesaplayabileceğiniz tespit sisteminin gerçek pozitif oranıdır. Bu, sisteminizin algılayabildiği kötü amaçlı yazılım örneklerinin yüzdesini hesapladığından, sisteminizin kötü amaçlı yazılım "gördüğünde" kötü amaçlı yazılımı tanıma yeteneğini ölçer.

Ancak, tespit sisteminizin kötü amaçlı yazılım gördüğünde alarm vereceğini bilmek, doğruluğunu değerlendirmek için yetersizdir. Örneğin, bir değerlendirme kriteri olarak yalnızca gerçek pozitif oranı kullandıysanız, tüm dosyalarda "evet, bu kötü amaçlı yazılımdır" diyen basit bir işlev, mükemmel bir gerçek pozitif oran verecektir. Bir tespit sisteminin gerçek testi, kötü amaçlı yazılım gördüğünde "evet, bu kötü amaçlı yazılımdır" ve zararlı yazılım gördüğünde "hayır, bu kötü amaçlı yazılım değildir" demesidir.

Bir sistemin kötü amaçlı yazılım olup olmadığını ayırt etme yeteneğini ölçmek için, sistemin yanlış pozitif oranını da ölçmeniz gerekir; bu, sisteminizin iyi huylu yazılım gördüğünde kötü amaçlı yazılım alarmı verme hızıdır. Sisteminizin yanlış pozitif oranını, sistemin kötü amaçlı yazılım olarak işaretlediği iyi huylu örneklerin sayısını, test edilen iyi huylu örneklerin toplam sayısına bölerek hesaplayabilirsiniz.

->Relationship Between True and False Positive Rates:Doğru ve Yanlış Pozitif Oranlar Arasındaki İlişki

Bir algılama sistemi tasarlarken, gerçek pozitif oranı olabildiğince yüksek tutarken yanlış pozitif oranını olabildiğince düşük tutmak istersiniz. Her zaman doğru olan gerçekten mükemmel bir kötü amaçlı yazılım algılama sistemi oluşturmazsanız (bu, kötü amaçlı yazılımın gelişen doğası göz önüne alındığında gerçekten imkansızdır), yüksek bir gerçek pozitif arzusu ile düşük bir yanlış pozitif oranı arzusu arasında her zaman gerilim olacaktır.

Durumun neden böyle olduğunu görmek için, bir ikili dosyanın kötü amaçlı yazılım olup olmadığına karar vermeden önce, ikili için bir şüpheli puan oluşturmak için ikili dosyanın kötü amaçlı yazılım olduğuna dair tüm kanıtları toplayan bir algılama sistemi hayal edin. Bu varsayımsal şüphecilik-puan oluşturma sistemine MalDetect diyelim. Şekil 7-2, MalDetect'in, dairelerin ayrı yazılım ikili dosyalarını temsil ettiği 12 örnek ikili için çıkarabileceği değerlerin bir örneğini gösterir. Bir ikili dosya ne kadar sağa doğru, MalDetect tarafından verilen şüpheli puan o kadar yüksek olur.

Şekil 7-2: Ayrı yazılım ikili dosyaları için varsayımsal MalDetect sistemi tarafından şüpheli puanların çıktısı

threshold eşik

Şüpheli puanlar bilgilendiricidir, ancak MalDetect'in dosyalarımızdaki gerçek pozitif oranını ve yanlış pozitif oranını hesaplamak için, MalDetect'in şüpheli puanlarını, belirli bir yazılım ikili dosyasının kötü amaçlı olup olmadığına ilişkin olarak "evet" veya "hayır" yanıtlarına dönüştürmemiz gerekir. Bunu yapmak için bir eşik kuralı kullanıyoruz. Örneğin, şüpheli puanın bir sayıdan büyük veya bir sayıya eşit olması durumunda, söz konusu ikili dosyanın kötü amaçlı yazılım alarmı oluşturacağına karar veririz. Puan eşikten düşükse, eşikten düşük değildir.

Böyle bir eşik kuralı, bir şüphecilik puanını ikili tespit seçimine dönüştürmenin standart yoludur, ancak eşiği nerede belirlemeliyiz? Sorun şu ki doğru cevap yok. Şekil 7-3 conun-drum'ı gösterir: eşiği ne kadar yükseğe ayarlarsak, yanlış pozitif alma olasılığımız o kadar az olur, ancak yanlış negatifler alma olasılığımız o kadar artar.

Şekil 7-3: Bir eşik değerine karar verirken yanlış pozitif oranı ile gerçek pozitif oran arasındaki ilişkinin bir örneği

Örneğin, Şekil 7-3'te gösterilen en soldaki eşiği düşünelim, burada eşiğin solundaki ikili dosyalar iyi huylu olarak ve sağındaki ikili dosyalar kötü amaçlı yazılım olarak sınıflandırılır. Bu eşik düşük olduğu için, büyük bir gerçek pozitif oran (kötü amaçlı yazılım örneklerinin yüzde 100'ünü doğru bir şekilde sınıflandırarak), ancak korkunç bir yanlış pozitif oranı elde ediyoruz (iyi huylu örneklerin yüzde 33'ünü yanlışlıkla kötü niyetli olarak sınıflandırıyoruz).

Sezgimiz, eşiği yükseltmek olabilir, böylece yalnızca daha yüksek şüpheli puanlara sahip numuneler kötü amaçlı yazılım olarak kabul edilir. Böyle bir çözüm Şekil 7-3'te orta eşikte verilmiştir. Burada yanlış pozitif oranı 0.17'ye düşüyor, ancak maalesef gerçek pozitif oran da 0.83'e düşüyor. En sağdaki eşikte gösterildiği gibi eşiği sağa kaydırmaya devam edersek, yanlış pozitifleri ortadan kaldırır, ancak kötü amaçlı yazılımın yalnızca yüzde 50'sini tespit ederiz.

Gördüğünüz gibi mükemmel eşik diye bir şey yoktur. Düşük bir yanlış pozitif oranı (iyi) veren bir algılama eşiği, daha fazla kötü amaçlı yazılımı gözden kaçırma eğiliminde olacak ve düşük bir gerçek pozitif oranı (kötü) verecektir. Tersine, yüksek bir gerçek pozitif orana (iyi) sahip bir algılama eşiğinin kullanılması da yanlış pozitif oranını (kötü) artıracaktır.

->ROC Curves:ROC Eğrileri

Algılama sistemlerinin gerçek pozitif oranı ile yanlış pozitif oranı arasındaki değiş tokuş, yalnızca kötü amaçlı yazılım algılayıcıları değil, tüm algılayıcılar için evrensel bir sorundur. Mühendisler ve istatistikçiler bu fenomen hakkında uzun uzun düşündüler ve onu tanımlamak ve analiz etmek için Alıcı Çalışma Karakteristiği (ROC)Receiver Operating Characteristic eğrisini buldular.

Not:Alıcı Çalışma Karakteristiği ifadesiyle kafanız karıştıysa, endişelenmeyin - bu ifade kafa karıştırıcıdır ve fiziksel nesnelerin radar tabanlı tespiti olan ROC eğrilerinin orijinal olarak geliştirildiği bağlamla ilgilidir.

ROC eğrileri, çeşitli eşik ayarlarında ilişkili gerçek pozitif oranlarına karşı yanlış pozitif oranları çizerek bir algılama sistemini karakterize eder. Bu, daha düşük yanlış pozitif oranlar ile daha yüksek gerçek pozitif oranlar arasındaki ödünleşimi değerlendirmemize ve bunu yaparken durumumuz için "en iyi" eşiği belirlememize yardımcı olur.

Örneğin, Şekil 7-3'teki varsayımsal MalDetect sistemimiz için, sistemin gerçek pozitif oranı, yanlış pozitif oranı 0 (düşük eşik) olduğunda 05 ve yanlış pozitif oranı 033 (yüksek eşik değeri) olduğunda sistemin gerçek pozitif oranı 100'dür. ).

Şekil 7-4 bunun nasıl çalıştığını daha ayrıntılı olarak göstermektedir.

Şekil 7-4: ROC eğrilerinin ne anlama geldiğini ve nasıl oluşturulduklarını gösteren bir örnek

ROC eğrisini oluşturmak için, Şekil 7-3'te kullanılan üç eşikle başlıyoruz ve bunların sonuçtaki yanlış ve gerçek pozitif oranlarını Şekil 7-3'ün sol yarısında gösterildiği gibi çiziyoruz. Şekil 7-4'ün sağındaki grafik aynı şeyi gösterir, ancak tüm olası eşikler için. Gördüğünüz gibi, yanlış pozitif oranları ne kadar yüksekse, gerçek pozitif oranlar o kadar yüksek olur. Benzer şekilde, yanlış pozitif oranları ne kadar düşükse, gerçek pozitif oranlar o kadar düşük olur.

ROC eğrisinin "eğrisi curve", algılama sistemimizin tüm olası yanlış pozitif değerler üzerinde gerçek pozitif oranında nasıl yapacağını düşündüğümüzü ve algılama sistemimizin onun üzerinde nasıl yapacağını düşündüğümüzü temsil eden, iki boyutlu ROC grafiği içindeki bir çizgidir. tüm olası gerçek pozitif değerler üzerinden yanlış pozitif oranı. Böyle bir eğri oluşturmanın birçok yolu vardır, ancak bu, bu kitabın kapsamını aşmaktadır.

Bununla birlikte, basit bir yöntem, birçok eşik değerini denemek, karşılık gelen yanlış ve doğru pozitif oranları gözlemlemek, bunları çizmek ve noktaları bir çizgi kullanarak birleştirmektir. Şekil 7-4'ün sağ grafiğinde gösterilen bu bağlantılı çizgi bizim ROC eğrimiz olur.

->Considering Base Rates in Your Evaluation:Değerlendirmenizde Taban Oranları Dikkate Alma

Gördüğünüz gibi, ROC eğrileri, sisteminizin kötü niyetli ikili dosyaları kötü amaçlı olarak çağırma oranı (gerçek pozitif oran) ve zararsız ikili dosyaları kötü niyetli olarak adlandırdığı oran (yanlış pozitif oran) açısından nasıl performans göstereceğini size söyleyebilir. Ancak, ROC eğrileri size sistemin hassasiyeti dediğimiz gerçek pozitifler olacak sistem alarmlarının yüzdesini söylemeyecektir. Bir sistemin hassasiyeti, temel oran dediğimiz, aslında kötü amaçlı yazılım olan sistemin karşılaştığı ikili dosyaların yüzdesi ile ilgilidir. İşte her terimin dökümü:

Precision:Kesinlik Gerçek pozitif olan sistem algılama alarmlarının yüzdesi (yani gerçek kötü amaçlı yazılım algılamalarıdır). Başka bir deyişle, kesinlik, bazı ikili dosyalara karşı test edildiğinde algılama sisteminin gerçek pozitif / (doğru pozitifler + yanlış pozitifler) sayısıdır.

Base rate:Temel oran Aradığımız kaliteye sahip sisteme beslenen verilerin yüzdesi. Bizim durumumuzda taban oran, aslında kötü amaçlı yazılım olan ikili dosyaların yüzdesini ifade eder.

Bir sonraki bölümde bu iki ölçümün nasıl ilişkili olduğunu tartışacağız.

->How Base Rate Affects Precision:Taban Oran, Kesinliği Nasıl Etkiler?

Bir tespit sisteminin doğru ve yanlış pozitif oranları, taban oran değiştiğinde değişmese de, sistemin hassasiyeti kötü amaçlı yazılım taban oranındaki değişikliklerden etkilenir - çoğu zaman önemli ölçüde. Bunun neden doğru olduğunu görmek için aşağıdaki iki durumu ele alalım.

MalDetect'in yanlış pozitif oranının yüzde 1 ve gerçek pozitif oranın yüzde 100 olduğunu varsayalım. Şimdi, MalDetect'i önceden üzerinde kötü amaçlı yazılım olmadığını bildiğimiz bir ağda serbest bıraktığımızı varsayalım (belki ağ bir laboratuvarda sıfırdan oluşturulmuştur). Ağda kötü amaçlı yazılım olmadığını önceden bildiğimiz için, MalDetect'in attığı her alarm tanımı gereği yanlış bir pozitif olacaktır, çünkü MalDetect'in karşılaştığı tek ikili iyi huylu yazılım olacaktır. Başka bir deyişle, hassasiyet yüzde 0 olacaktır.

Buna karşılık, MalDetect'i tamamen kötü amaçlı yazılımdan oluşan bir veri kümesinde çalıştırırsak, alarmlarından hiçbiri yanlış pozitif olmayacaktır: Yazılım veri kümesinde hiçbir zararsız yazılım olmadığından, MalDetect'in yanlış bir pozitif üretme fırsatı asla olmayacaktır. Bu nedenle hassasiyet yüzde 100 olacaktır.

Bu uç durumların her ikisinde de, taban oranların MalDetect'in hassasiyeti veya alarmının yanlış pozitif olma olasılığı üzerinde büyük etkisi vardır.

->Estimating Precision in a Deployment Environment:Bir Dağıtım Ortamında Hassasiyeti Tahmin Etme

Artık bir test veri kümesindeki (taban oran) kötü amaçlı yazılım oranına bağlı olarak, sisteminizin çok farklı hassasiyet değerleri vereceğini biliyorsunuz. Sisteminizin sahip olacağı kesinliği, onu yerleştirdiğiniz ortamın taban oranının bir tahminine dayanarak tahmin etmek isterseniz ne olur? Tek yapmanız gereken, kesinlik formülündeki değişkenleri tahmin etmek için dağıtım ortamınızın tahmini taban oranını kullanmaktır: gerçek pozitifler / (doğru pozitifler + yanlış pozitifler). Üç numaraya ihtiyacınız olacak:

True positive rate (TPR):Sistemin gerçek pozitif oranı (TPR) veya sistemin doğru bir şekilde algılayacağı kötü amaçlı yazılım örneklerinin yüzdesi

False positive rate(FPR):Sistemin yanlış pozitif oranı (FPR) veya sistemin hatalı olarak alarm vereceği iyi huylu örneklerin yüzdesi

Base rate(BR):Sistemi kullanacağınız ikili dosyaların taban oranı (BR) (örneğin, sisteminizi kullanacağınız şey buysa, korsan sitelerden indirilmesini beklediğiniz ikili dosyaların yüzdesi kötü amaçlı yazılım olacaktır)

Kesinlik denkleminin payı - gerçek pozitif sayısı - gerçek pozitif oran × taban oran ile tahmin edilebilir ve size sisteminizin doğru şekilde tespit edeceği kötü amaçlı yazılım yüzdesini verir. Benzer şekilde, denklemin paydası - yani (gerçek pozitifler + yanlış pozitifler) - gerçek pozitif oran × taban oran + yanlış pozitif oran × (1 -taban oran) ile tahmin edilebilir ve size sistemdeki tüm ikili değerlerin yüzdesini verir Doğru tespit edilecek kötü amaçlı yazılım ikili dosyalarının sayısı artı yanlış pozitiflerin verileceği iyi huylu yazılım ikili dosyalarının fraksiyonunu hesaplayarak alarm verecektir.

Özetle, sisteminizin beklenen hassasiyetini şu şekilde hesaplarsınız:

Taban oranın bir algılama sisteminin performansı üzerinde nasıl derin bir etkisi olabileceğini görmek için başka bir örnek ele alalım. Örneğin, yüzde 80 gerçek pozitif oranı ve yüzde 10 yanlış pozitif oranı olan bir algılama sistemimiz olduğunu ve onu çalıştırdığımız yazılım ikili dosyalarının yüzde 50'sinin kötü amaçlı yazılım olması beklendiğini varsayalım. Bu, yüzde 89'luk beklenen bir hassasiyete yol açacaktır. Ancak taban oran yüzde 10 olduğunda, hassasiyetimiz yüzde 47'ye düşer. Taban oranımız çok düşükse ne olur? Örneğin, modern bir kurumsal ağda, çok az yazılım ikili dosyası aslında kötü amaçlı yazılımdır. Hassasiyet denklemimizi kullanarak, yüzde 1'lik bir taban oranı varsayarsak (100 ikili dosyadan 1'i kötü amaçlı yazılımdır), yaklaşık yüzde 7,5'lik bir hassasiyet elde ederiz, bu da sistemimizin alarmlarının yüzde 92,5'inin yanlış pozitif olacağı anlamına gelir!

Ve yüzde 0,1'lik bir taban oranı varsayarsak (1000 ikili dosyadan 1'inin kötü amaçlı yazılım olma olasılığı yüksektir), yüzde 1'lik bir hassasiyet elde ederiz, bu da sistem alarmlarımızın yüzde 99'unun yanlış pozitifler olacağı anlamına gelir! Son olarak, yüzde 0,01'lik bir taban oranda (10.000 ikili dosyadan 1'inin kötü amaçlı yazılım olma olasılığı yüksektir - muhtemelen bir kurumsal ağdaki en gerçekçi varsayımdır), beklenen hassasiyetimiz yüzde 0,1'e düşer, bu da sistemimizin uyarılarının ezici çoğunluğunun yanlış olacağı anlamına gelir pozitifler. Bu analizden bir çıkarım, yüksek yanlış pozitif oranlarına sahip algılama sistemlerinin kurumsal ortamlarda neredeyse hiçbir zaman yararlı olmayacağıdır, çünkü hassasiyetleri çok düşük olacaktır. Bu nedenle, kötü amaçlı yazılım tespit sistemleri oluşturmanın temel amaçlarından biri, yanlış pozitif oranı sistem hassasiyetinin makul olacağı şekilde en aza indirmektir.

Bir diğer ilgili çıkarım, bu bölümde daha önce tanıtılan ROC eğrisi analizini yaptığınızda, sisteminizi bir kuruluş ortamında dağıtılacak şekilde geliştiriyorsanız, örneğin yüzde 1'in üzerindeki yanlış pozitif oranları etkili bir şekilde göz ardı etmeniz gerektiğidir, çünkü daha yüksek yanlışlar pozitif oran, muhtemelen işe yaramaz hale getirilecek kadar düşük hassasiyete sahip bir sistemle sonuçlanacaktır.

->ÖZET:

Bu bölümde, gerçek pozitif oran, yanlış pozitif oran, ROC eğrileri, taban oranları ve kesinlik dahil olmak üzere temel algılama değerlendirme kavramlarını öğrendiniz. Gerçek pozitif oranı en üst düzeye çıkarmanın ve yanlış pozitif oranını en aza indirmenin, kötü amaçlı yazılım tespit sistemi oluşturmada ne kadar önemli olduğunu gördünüz. Taban oranının hassasiyeti etkileme şekli nedeniyle, algılama sisteminizi bir kuruluş içinde dağıtmak istiyorsanız, yanlış pozitif oranını azaltmak özellikle önemlidir.

Bu kavramlarda tamamen akıcı hissetmiyorsanız endişelenmeyin. Kötü amaçlı yazılım tespit sistemini sıfırdan oluşturup değerlendireceğiniz sonraki bölümde onlarla daha fazla pratik yapacaksınız. Süreçte, makine öğrenimine dayalı algılayıcılarınızı geliştirmenize yardımcı olacak ek makine öğrenimine özgü değerlendirme kavramlarını öğreneceksiniz.