تعریف ناهنجاری های شبکه

ناهنجاری های شبکه به هرگونه انحراف از رفتار عادی ترافیک شبکه اشاره دارد. این انحرافات می تواند نشان دهنده مسائل مختلفی مانند تهدیدات امنیتی، خرابی شبکه یا پیکربندی نادرست باشد. ناهنجاری‌ها می‌توانند به روش‌های مختلفی ظاهر شوند، از جمله الگوهای ترافیک غیرمعمول، افزایش غیرمنتظره استفاده، تلاش‌های دسترسی غیرمجاز و موارد دیگر.

انواع ناهنجاری های شبکه

انواع مختلفی از ناهنجاری های شبکه وجود دارد که هر کدام دارای ویژگی های منحصر به فرد و پیامدهای بالقوه است:

اوج ترافیک:

توضیحات: افزایش ناگهانی ترافیک شبکه که بیشتر از الگوهای استفاده عادی است.

دلایل: ممکن است به دلایل قانونی مانند به روز رسانی نرم افزار یا فعالیت های مخرب مانند حملات DDoS (Distributed Denial-of-Service) باشد.

مثال: یک وب‌سایت تجارت الکترونیک که در طول فروش تعطیلات، ترافیک ناگهانی را تجربه می‌کند.

دسترسی غیرمجاز:

توضیحات: تلاش برای دسترسی به منابع شبکه بدون مجوز مناسب.

دلایل: ممکن است به دلیل تلاش هکرها برای نفوذ به شبکه، دسترسی کارکنان به مناطق محدود شده یا حساب های کاربری در معرض خطر باشد.

مثال: یک دستگاه ناشناخته که تلاش می کند به شبکه داخلی شرکت متصل شود.

حملات DDoS:

توضیحات: حملات هماهنگ با هدف غلبه بر منابع شبکه از طریق پر کردن حجم عظیمی از ترافیک از منابع مختلف.

علل: معمولاً توسط باتنت‌های کنترل شده توسط مهاجمان آغاز می‌شود.

مثال: وب سایت یک موسسه مالی به دلیل حمله DDoS در مقیاس بزرگ در دسترس نیست.

الگوهای بسته غیرعادی:

توضیحات: الگوها یا پروتکل های بسته غیرمعمول که با رفتار شبکه مورد انتظار مطابقت ندارند.

علل: می تواند نشانه ای از اسکن شبکه، استخراج داده ها یا فعالیت بدافزار باشد.

مثال: افزایش تعداد بسته های ارسال شده به پورت های غیر استاندارد.

گرازهای پهنای باند:

توضیحات: دستگاه ها یا برنامه هایی که پهنای باند غیرمعمول زیادی مصرف می کنند.

دلایل: ممکن است برنامه‌های کاربردی با پهنای باند بالا یا فعالیت‌های مخرب مانند سرقت داده یا اشتراک‌گذاری غیرمجاز فایل باشد.

مثال: کامپیوتر یک کاربر که به طور مداوم فایل های بزرگ را دانلود می کند، بر عملکرد کلی شبکه تأثیر می گذارد.

اهمیت تشخیص ناهنجاری ها

تشخیص ناهنجاری های شبکه به چند دلیل حیاتی است:

امنیت:

تشخیص تهدید: شناسایی ناهنجاری‌ها می‌تواند به شناسایی تهدیدات امنیتی مانند نفوذ، آلودگی‌های بدافزار و نقض داده‌ها کمک کند.

جلوگیری از آسیب: تشخیص زودهنگام امکان واکنش سریع را فراهم می کند و به طور بالقوه از آسیب قابل توجهی به زیرساخت شبکه و از دست دادن داده ها جلوگیری می کند.

عملکرد شبکه:

بهینه سازی منابع: با شناسایی و رفع ناهنجاری ها، مدیران شبکه می توانند استفاده بهینه از منابع شبکه را تضمین کنند.

اجتناب از خرابی: تشخیص زودهنگام مشکلاتی مانند افزایش ترافیک یا حملات DDoS می‌تواند به کاهش تأثیر آنها، کاهش زمان خرابی و حفظ در دسترس بودن خدمات کمک کند.

انطباق:

الزامات نظارتی: بسیاری از صنایع دارای مقرراتی هستند که به نظارت و ایمن سازی ترافیک شبکه برای محافظت از داده های حساس نیاز دارند.

مسیرهای حسابرسی: شناسایی و ثبت ناهنجاری ها سابقه ای را ارائه می دهد که می تواند برای ممیزی ها و گزارش انطباق مفید باشد.

بهره وری عملیاتی:

مدیریت فعال: تشخیص ناهنجاری امکان مدیریت فعال شبکه را فراهم می کند و مسائل بالقوه را قبل از بحرانی شدن شناسایی می کند.

بهبود تجربه کاربری: اطمینان از اینکه شبکه بدون اختلال کار می کند، تجربه کلی کاربر را برای کارمندان و مشتریان بهبود می بخشد.

صرفه جویی در هزینه:

جلوگیری از نقض داده ها: سازمان ها با تشخیص زودهنگام و پاسخ به ناهنجاری ها می توانند از هزینه های بالای مربوط به نقض داده ها و حملات سایبری جلوگیری کنند.

کاهش زمان از کار افتادن شبکه: حفظ آپتایم شبکه، ضررهای مالی احتمالی مرتبط با اختلالات سرویس را کاهش می دهد.

الگوریتم جنگل جداسازی توضیح داده شد

Isolation Forest یک الگوریتم یادگیری ماشینی است که برای تشخیص ناهنجاری استفاده می شود. این کار با جداسازی ناهنجاری ها (پرت ها) از نقاط داده عادی از طریق یک سری تقسیمات تصادفی کار می کند. در اینجا یک توضیح آسان با یک مثال آورده شده است:

جنگل انزوا چگونه کار می کند

تقسیم تصادفی: این الگوریتم با انتخاب تصادفی یک ویژگی و یک مقدار تقسیم، یک سری درخت تصمیم ایجاد می کند. این فرآیند تا زمانی ادامه می یابد که هر نقطه داده ایزوله شود.

جداسازی: ناهنجاری ها سریعتر از نقاط عادی جدا می شوند زیرا دارای ویژگی های منحصر به فردی هستند که جداسازی آنها را آسان تر می کند.

امتیازدهی: تعداد تقسیم‌های مورد نیاز برای جداسازی یک نقطه یک امتیاز ناهنجاری می‌دهد. نقاطی که به شکاف کمتری نیاز دارند، احتمالاً ناهنجاری هستند.

یادگیری تحت نظارت

تعریف: یادگیری نظارت شده نوعی از یادگیری ماشینی است که در آن مدل بر روی داده های برچسب دار آموزش داده می شود. این بدان معنی است که داده ها با جفت های ورودی–خروجی می آیند، جایی که خروجی (برچسب) از قبل مشخص است.

چگونه کار می کند: مدل یاد می گیرد که بر اساس داده های ورودی با مقایسه پیش بینی های خود با برچسب های شناخته شده در طول آموزش، پیش بینی یا تصمیم گیری کند. پارامترهای خود را تنظیم می کند تا تفاوت بین پیش بینی های خود و برچسب های واقعی را به حداقل برساند.

مثال: تصور کنید مجموعه داده‌ای از خانه‌ها با ویژگی‌هایی مانند اندازه، تعداد اتاق‌ها و موقعیت مکانی دارید و هر خانه دارای قیمتی است. با آموزش یک مدل یادگیری نظارت شده بر روی این داده ها، می توانید قیمت یک خانه جدید را بر اساس ویژگی های آن پیش بینی کنید.

یادگیری بدون نظارت

تعریف: یادگیری بدون نظارت نوعی از یادگیری ماشینی است که در آن مدل بر روی داده های بدون برچسب آموزش داده می شود. هدف، کشف الگوها یا ساختارهای پنهان در داده ها است.

چگونه کار می‌کند: مدل سعی می‌کند ساختار ذاتی داده‌ها را با یافتن شباهت‌ها یا تفاوت‌ها بیاموزد، بدون اینکه اطلاعات قبلی در مورد خروجی باید داشته باشد.

مثال: اگر مجموعه داده‌ای از داده‌های مشتری با ویژگی‌هایی مانند سن، درآمد و عادات هزینه دارید، یک الگوریتم یادگیری بدون نظارت می‌تواند مشتریان را بر اساس ویژگی‌های مشابه به خوشه‌هایی گروه‌بندی کند که می‌تواند برای تقسیم‌بندی بازار مفید باشد.

یادگیری تقویتی

تعریف: یادگیری تقویتی نوعی از یادگیری ماشینی است که در آن یک عامل یاد می گیرد با تعامل با یک محیط تصمیم گیری کند. نماینده بر اساس اقدامات خود پاداش یا جریمه دریافت می کند و هدف آن به حداکثر رساندن پاداش تجمعی در طول زمان است.

چگونه کار می کند: عامل محیط را بررسی می کند و اقداماتی را انجام می دهد و بازخوردهایی را به شکل پاداش یا مجازات دریافت می کند. از این بازخورد برای یادگیری یک استراتژی (یا خط مشی) استفاده می کند که کل پاداش آن را به حداکثر می رساند.

مثال: روباتی را در نظر بگیرید که در حال یادگیری مسیریابی در پیچ و خم است. هر بار که ربات به انتهای پیچ و خم می رسد، یک جایزه دریافت می کند. اگر به دیوار برخورد کند، جریمه می شود. با گذشت زمان، ربات بهترین مسیر را برای عبور سریع و کارآمد از پیچ و خم می آموزد.

خوشه بندی

خوشه بندی تکنیکی است که نقاط داده مشابه را با هم گروه بندی می کند. در تشخیص ناهنجاری، الگوریتم های خوشه بندی می توانند به شناسایی نقاط پرت کمک کنند، که نقاط داده ای هستند که به خوبی در هیچ خوشه ای قرار نمی گیرند. الگوریتم های خوشه بندی رایج عبارتند از:

K-Means Clustering: این الگوریتم داده ها را به پارتیشن بندی می کند

k خوشه، که در آن هر نقطه داده متعلق به خوشه با نزدیکترین میانگین است. نقاط پرت نقاط داده ای هستند که از هر مرکز خوشه ای دور هستند.

DBSCAN (خوشه‌بندی فضایی برنامه‌های کاربردی با نویز مبتنی بر چگالی): این الگوریتم نقاط داده‌ای را که به طور نزدیک با هم بسته‌بندی شده‌اند را گروه‌بندی می‌کند و نقاطی را که در مناطق با چگالی کم (یعنی دور از هر خوشه‌ای) هستند به عنوان نقاط پرت علامت‌گذاری می‌کند.

مثال: تصور کنید مجموعه داده ای از الگوهای ترافیک شبکه دارید. خوشه‌بندی می‌تواند الگوهای مشابه را با هم گروه‌بندی کند و هر الگوی که در این گروه‌ها قرار نگیرد، می‌تواند به‌عنوان غیرعادی علامت‌گذاری شود.

طبقه بندی

طبقه بندی شامل پیش بینی دسته ای است که یک نقطه داده جدید بر اساس مجموعه داده آموزشی با برچسب های شناخته شده به آن تعلق دارد. برای تشخیص ناهنجاری، طبقه بندی کننده آموزش داده شده است تا بین داده های عادی و غیرعادی تمایز قائل شود.

ماشین‌های بردار پشتیبانی (SVM): این الگوریتم ابرصفحه‌ای را پیدا می‌کند که به بهترین وجه داده‌ها را به کلاس‌های مختلف جدا می‌کند. در تشخیص ناهنجاری، می توان از آن برای جداسازی داده های عادی از ناهنجاری ها استفاده کرد.

شبکه‌های عصبی: اینها مجموعه‌ای از الگوریتم‌هایی هستند که به‌طور آزاد از مغز انسان مدل‌سازی شده‌اند و برای تشخیص الگوها طراحی شده‌اند. در تشخیص ناهنجاری، شبکه های عصبی می توانند یاد بگیرند که نقاط داده را به عنوان عادی یا غیرعادی طبقه بندی کنند.

مثال: در تشخیص تقلب، یک الگوریتم طبقه‌بندی می‌تواند از داده‌های تراکنش تاریخی با برچسب «متقلب» یا «غیر متقلبانه» یاد بگیرد. سپس تراکنش های جدید را می توان بر اساس این مدل آموخته شده طبقه بندی کرد.