تاریخ: .....
شماره: .....

عنوان پروژه: سیستم تشخیص ناهنجاری های شبکه با استفاده از یادگیری بدون نظارت
استاد راهنمای پروژه: دکتر مریم امیرحائری **امضاء:**
مشخصات دانشجو: **نام و نام خانوادگی: محمدمهدی آقاجانی** **گرایش: نرم افزار** **شماره دانشجوئی: ۹۳۳۱۰۵۶** **ترم ثبت نام پروژه: دوم ۹۶-۹۷**
داوران پروژه: **۱-** **امضاء داور:** **۲-** **امضاء داور:**
شرح پروژه (در صورت مشترک بودن بخشی از کار که بعهده دانشجو می باشد مشخص شود): هدف از این پروژه ارائه یک سیستم تشخیص ناهنجاری های ترافیک شبکه های کامپیوتری است. در این پروژه سیستمی طراحی و پیاده سازی خواهد شد که با دریافت داده های یک بازه زمانی از ترافیک یک شبکه، ناهنجاری‌های ترافیک شبکه را تشخیص می دهد. محصول نهایی را به صورت یک وب سایت ارایه می‌شود که کاربر آن می‌تواند یک داده ترافیک شبکه بارگذاری نماید و ناهنجاری هایی که متوجه آن شبکه هست را به طور تحلیلی مشاهده نماید. روش تشخیص این ناهنجاری‌ها به صورت بدون ناظر خواهد بود این روشها این امکان را فراهم می‌کنند که بدون دانش اولیه از ناهنجاری‌های شبکه و داده های برچسب دار امکان تشخیص ناهنجاری را فراهم می‌آورد.
وسائل مورد نیاز:
محل انجام پروژه: دانشگاه صنعتی امیرکبیر **تاریخ شروع:**

این قسمت توسط دانشکده تکمیل میگردد:

تاریخ تصویب در گروه: **اسم و امضاء:** **تاریخ تصویب در دانشکده:** **اسم و امضاء:** **اصلاحات لازم در تعریف پروژه:**

توجه: پروژه حداکثر یک ماه و نیم پس از شروع ترمی که در آن در درس پروژه ثبت نام بعمل آمده است باید به تصویب برسد.

نسخه۱- دانشکده	نسخه ۲- استاد راهنما	نسخه ۳- دانشجو

تعریف پروژه

امروزه با رشد روزافزون تکنولوژی و کاربرد شبکه در سازمان ها و رشد استفاده از اینترنت ، به طور همزمان میزان تهدیدات امنیتی آن ها نیز افزایش یافته است در نتیجه اهمیت تشخیص تهدیدات و حملات بیش از پیش خودنمایی میکند. تشخیص این حملات و تهدیدات میتواند با شیوه های مختلفی انجام شود اما به کارگیری متد های نوین علمی به خصوص هوش مصنوعی در راه تشخیص این گونه تهدیدات میتواند کارایی مناسبی را حاصل نماید.

در این پروژه قصد داریم تا با استفاده از داده های ترافیک یک شبکه که از پیش آن را آماده کرده ایم یک روش برای بررسی ترافیک آن شبکه و تشخیص ناهنجاری ها در آن شبکه ارایه نماییم. این روش ترافیک شبکه را در سطح بسته[1] ها بررسی می نماید. همچنین بررسی ها به طور آفلاین انجام خواهد شد. متد استفاده شده از مدل های آماری و هم چنین روش های خوشه بندی[2] استفاده خواهد نمود.

روند کار بدین صورت است که ابتدا نیازمندی های پروژه را تحلیل کرده و بعد یک الگوریتم تشخیص داده پرت بدون نظارت بر روی داده ها اعمال میکنیم سپس الگوریتم پیاده سازی شده را تست و بررسی میکنیم و اگر خطا داشت آن را بهبود میدهیم.

محصول نهایی را به صورت یک وب سایت ارایه میکنیم که کاربر آن میتواند یک داده ترافیک شبکه بارگذاری نماید و ناهنجاری هایی که متوجه آن شبکه هست را به طور تحلیلی مشاهده نماید.

هدف پروژه

با توجه به مطالب ارایه شده در بالا ، هدف اصلی این پروژه یافتن الگوی های ناهنجار در داده های ترافیک شبکه می باشد که با استفاده از متد های تشخیص داده پرت[3] بدون نظارت قصد پیدا کردن این داده ها را داریم. خروجی این تحلیل ها موجب خواهد شد که بتوان حملات و تهدیدات امنیتی در شبکه را تشخیص داد.

بررسی کارهای انجام شده

امروزه سیستم های تشخیص نفوذ در شبکه بر اساس پایگاه داده ای از اطلاعات از پیش تعیین شده ای که باید دایما به روزرسانی شود ، کار میکنند. اما با استفاده از تکنیک های بدون نظارت تشخیص ناهنجاری میتوان این نفوذ ها را تشخیص داد. البته این سیستم ها بسیار پیچیده می باشند و همین امر باعث میشود که در عمل نیازمندی های انجام شده را برآورده نکنند. درومارد و همکاران [1] ابتدا روشی بدون نظارت به نام UNADA را برای کشف ناهنجاری ها در شبکه ارائه نموده‌اند و سپس با استفاده از الگوریتم ORUNADA مدل برخط آن را معرفی کرده اند [۲] .

[1] packet
[2] clustering
[3] Outlier detection

در روش UNADA ابتدا ترافیک شبکه در سطح جریان[1] را جمع آوری می کند. سپس فضای داده ها برای از بین بردن اثر نامطلوب بعد زیاد[2] تبدیل به زیر فضا می شوند. هر زیرفضا در واقع انتخاب دو ویژگی[3] از بردار ویژگی ها می باشد در نتیجه $N = \frac{n*(n-1)}{2}$ تا زیر فضا خواهیم داشت که در اینجا n طول بردار ویژگی ها می باشد. سپس بر روی هر زیرفضا الگوریتم DB Scan اجرا می شود و هر زیرفضا خوشه بندی می گردد. بعد از اجرای خوشه بندی برخی از جریان ها وارد هیچ خوشه ای نشده اند که به هر یک از این خوشه ها امتیازی اختصاص داده می شود و به هر جریانی که در زیرفضای مربوطه وارد یک خوشه شده است امتیاز صفر داده می شود. بعد بر اساس مجموع امتیازات جریان ها در تمامی زیرفضا ها امتیاز کلی آن ها در نظر گرفته می شود و هر جریانی که از حدی معلوم امتیاز بیشتری داشته باشد به عنوان ناهنجاری در نظر گرفته می شود.

متدلوژی پیاده سازی نرم افزار

در این پروژه از مدل فرآیند آبشاری استفاده میکنیم که به صورت ترتیبی هر فاز را به جلو می برد و با توجه به ماهیت سیستم پیش رو بیشترین تطبیق را با پروژه دارد. در این روش البته کمترین بازخورد بین فاز ها وجود دارد. اولین فاز در قسمت بالا سمت چپ شکل ۱ قرار گرفته و فاز های بعدی به ترتیب به دنبال آن مشخص شده اند. هر فاز در واقع یک محصول کاری می باشد که نتایج آن در قالب سند ها یا خروجی های قابل تحویل برای فاز بعدی مورد استفاده قرار می گیرند.

مدل آبشاری که گاهی مدل دوران زندگی کلاسیک نامیده می شود، روشی سیستماتیک و ترتیبی را طی توسعه نرم افزاری که از مشخصه نیاز های مشتری آغاز می شود و تا برنامه ریزی، مدل سازی ، ساخت و استقرار ادامه می یابد، پیشنهاد می کند. [2]

معماری سیستم

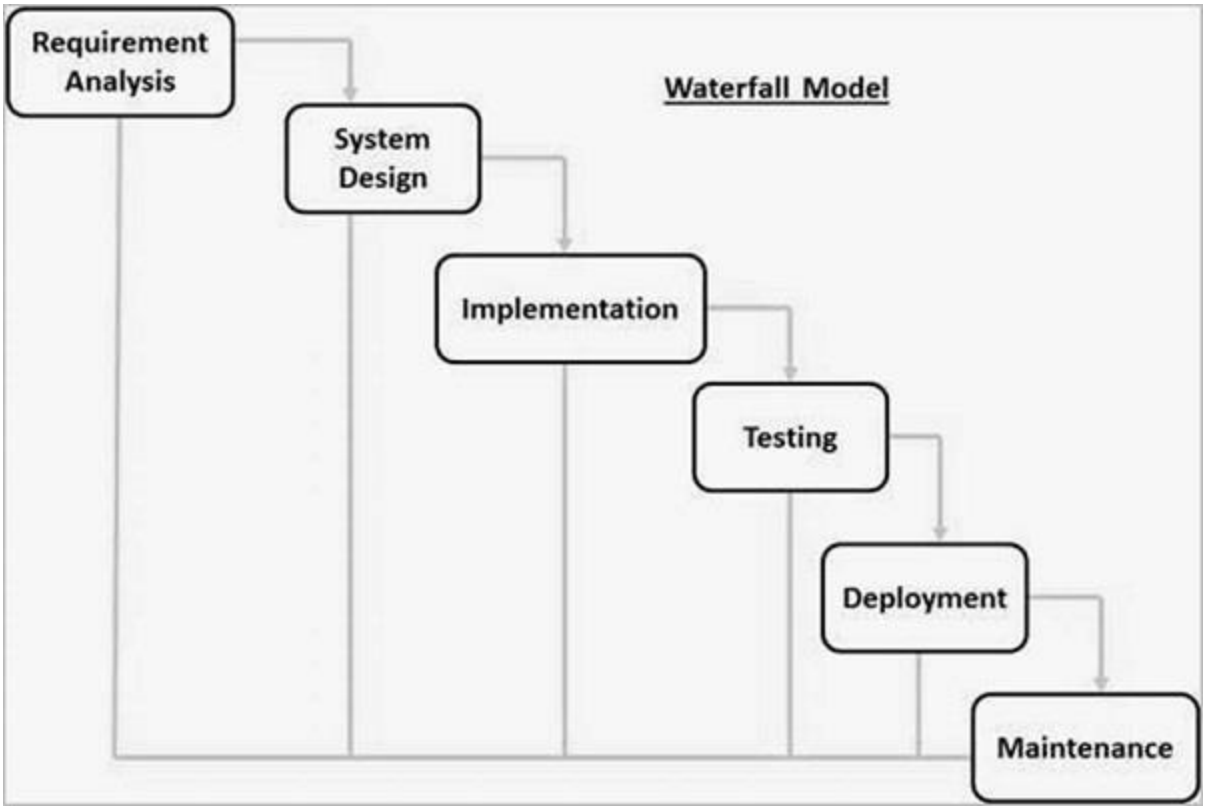
این سیستم دارای دو بخش یا سرویس کلی است:

- بخش تحلیل داده
- رابط کاربری وب سایت

در پیاده سازی وب سایت از MVC (Model-View-Controller) بهره خواهیم گرفت. استفاده از MVC به خاطر جدا سازی لایه های مختلف برنامه می باشد تا بتوان به هر یک به طور جداگانه و با دقت بیشتری پرداخت.

در بخش تحلیل داده با بهره گیری از زبان برنامه‌نویسی پایتون و استفاده از ابزارهای یادگیری ماشین مانند Scikit و numpy , panda به تحلیل داده ها میپردازیم. ابزار numpy و panda در نگه داری داده ها و

[1] flow
[2] High Dimensionality
[3] feature

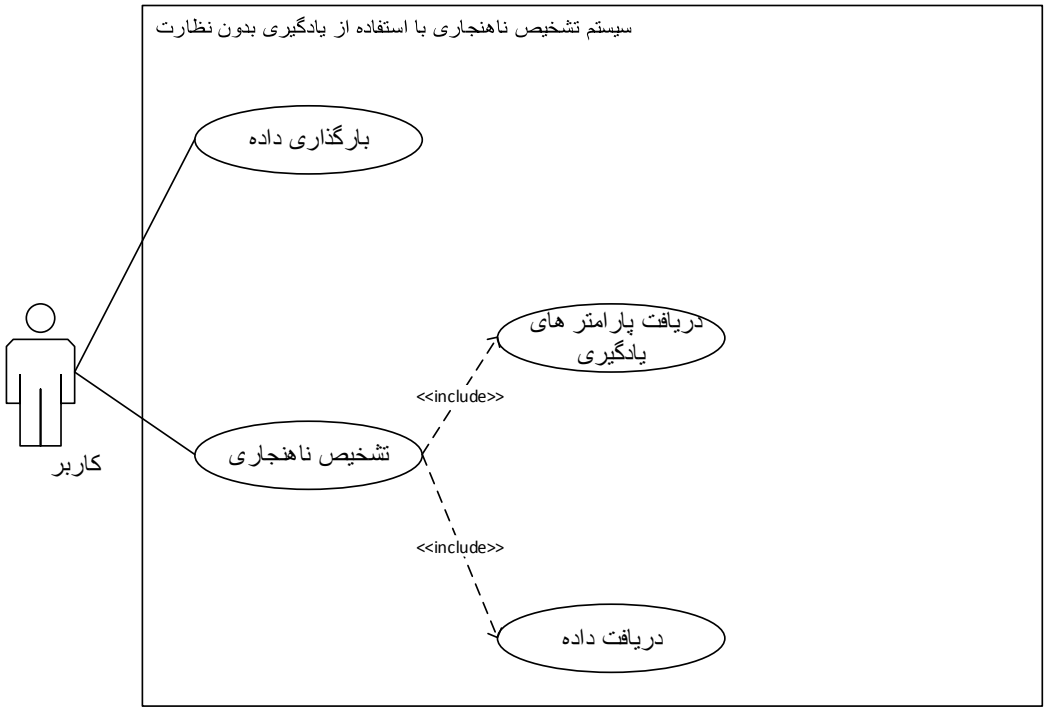


شکل ۱ متدولوژی آبشاری

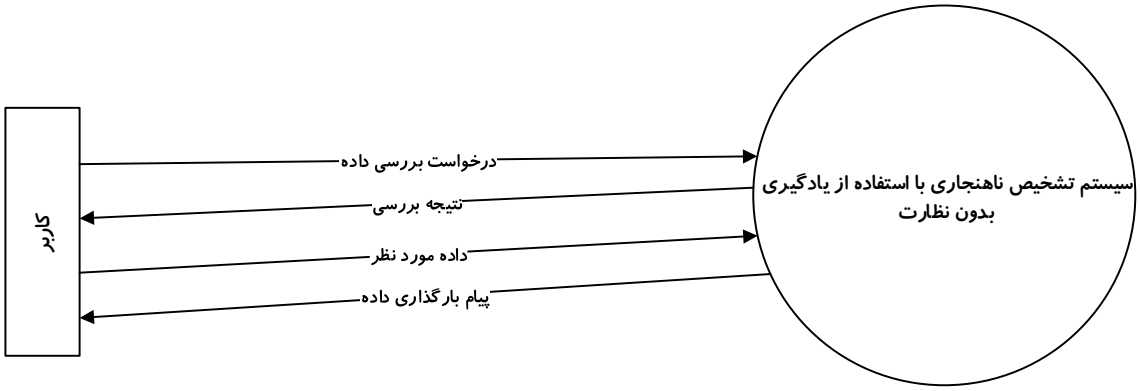
پردازش آن ها کمک می کنند و ابزار Scikit در واقع کتابخانه ای برای به کارگیری الگوریتم های رایج در یادگیری ماشین[1] می باشد.

از طرفی پایگاه داده ای وجود دارد که وب اپلیکیشن با آن در ارتباط است و داده های از پیش آماده شده شبکه بر روی آن وجود دارد. همچنین کاربرانی که داده ای بر روی سیستم بارگذاری میکنند نیز در این پایگاه داده ذخیره می شود. نمودار USE CASE سیستم ذکر شده در شکل ۲ دیده می شود. هم چنین نمودار context diagram نیز در شکل ۳ آورده شده است.

[1] Machine Learning



شکل ۲ نمودار use case



شکل ۳ نمودار context diagram

ابزارهای پیاده سازی

برای بخش تحلیل داده ها از زبان پایتون و از کتابخانه هایی که در بالا ذکر شد استفاده میکنیم. برای بخش وب سایت هم از JAVA در سمت سرور و به طور خاص از JSP استفاده می نماییم.

مراجع

[1] J. Dromard, G. Roudiére, and P. Owezarski. Online and scalable unsupervised network anomaly detection method. IEEE Transaction on Network and System Management (TNSM), 14(1), January 2016.

[2] R. Pressman, Software engineering. New York: Mcgraw-Hill, 2014.