PRACTICA DE MACHINE LEARNING

CASO DE USO DE MACHINE LEARNING EN LA CIBERSEGURIDAD

Autor: Fco Javier García Varela

Fecha de entrega: 27/11/2022

Descripción del caso de uso

¿Cuál es el problema?

Queremos crear un sistema que nos permita identificar cuales de las conexiones que recibimos hacia nuestros servidores son un ataque y cuales no antes de que el ataque tenga éxito y así poder evitarlo.

¿Cómo se está afrontando ahora?

Actualmente existen varios métodos que se emplean para la prevención de ataques, tales como firewalls, factores de segunda autentificación, antivirus, etc. Además, también muchos de ellos emplean diferentes métodos de Machine Learning para identificar cuales de las conexiones son legítimas y cuales no.

¿Acción que buscamos poder hacer para solucionar el problema?

Buscamos crear un nuevo método que mejore las predicciones de cuales de las conexiones entrantes en nuestros servidores se tratan de una amenaza o un posible ataque y cuales no, con el fin de automatizar alertas lo mas precias posibles.

KPIs – Indicadores de negocio

En nuestro caso trabajaremos con los siguientes KPIs:

- Porcentaje de amenazas detectadas
- Porcentaje de acierto en el tipo de ataque recibido

¿Cuáles son los mínimos que se esperan de este caso de uso?

Para este modelo esperamos obtener una tasa de aciertos lo suficientemente elevada para evitar que nuestros servidores se vean afectados, por lo que los valores esperados son los siguientes:

- Porcentaje de amenazas detectadas: 95%
- Porcentaje de acierto en el tipo de ataque recibido: 80%

Validación: ¿Qué criterio se va a usar para decidir si la solución es aceptable?

Para que el modelo se considere valido además de ser capaz de detectar la gran mayoría de las amenazas también es importante que el número de falsos positivos sea lo menor posible para evitar activar el equipo de ciberseguridad sin necesidad y para evitar que un usuario legitimo se vea afectado sin motivo, por lo que el porcentaje de falsos positivos aceptado es del 15% para que el modelo se considere viable.

Experimentación: ¿Cómo vamos a corroborar el funcionamiento?

Para comprobar el correcto funcionamiento del modelo realizaremos prueba monitorizadas donde veremos si el modelo es capaz de detectar ataques concretos.

Además, se realizarán pruebas periódicas para validar que el modelo sigue siendo efectivo a los nuevos tipos de ataques que se van descubriendo

Productivización: ¿Qué salida debe tener la solución que se desarrolle?

En caso de que el modelo sea viable y valido, se usara de forma conjunta con un firewall que analizara todo el trafico entrante en nuestros servidores y en caso de detectar alguna conexión no legitima creara un aviso en nuestro sistema de monitorización.

Equipo de trabajo

Para llevar a cabo este proyecto deberemos contar con los siguientes perfiles:

- Analista de Big Data
- Analista de ciberseguridad
- Programador de Python
- Pentester

Detalle del caso de uso

Detalle funcional

Dado que todos los servidores expuestos en internet son victimas de ataques, queremos lograr un sistema que detecte lo mas rápida y eficazmente los ataques para poder bloquearlos antes de que logren su objetivo.

Dado que trabajamos con datos de clientes debemos cumplir en todo momento la ley de protección de datos, tanto al trabajar con los datos de prueba para desarrollar el modelo como después cuando usemos el modelo en tiempo real para revisar los datos de las conexiones a nuestros servidores.

Actualmente tenemos diferentes reglas ya establecidas en nuestros firewalls que debemos tener en cuenta a la hora de desarrollar nuestro modelo, como los puertos permitidos para la conexión o las IP´s ya conocidas para ciertos servicios.

Identificación de orígenes de datos

Vamos a emplear para desarrollar nuestro modelo un dataset recogido de las conexiones a nuestros servidores durante el ultimo año. De ella sabemos cuales han sido

ataques y cuales son conexiones legitimas, por lo que emplearemos un modelo supervisado para desarrollar nuestro modelo.

En este dataset tendremos ya descritas las diferentes variables que usaremos para identificar todas y cada una de las diferentes conexiones y ver todas las similitudes y diferencias entre ellas con el fin de diferencias las conexiones legitimas de los ataques.

Desarrollo del caso de uso

Puntos intermedios o seguimiento

Durante el desarrollo de nuestro método se irán realizando diferentes comprobaciones para asegurarnos que nuestro modelo mantiene un equilibrio entre el Bias y la Variabilidad de nuestro modelo.

También se realizarán pruebas diferentes pruebas con datos ya analizados para verificar si las predicciones de nuestro modelo son correctas.

Como hemos dicho antes, nuestro modelo seguirá un esquema de aprendizaje supervisado, ya que disponemos de los datos ya clasificados para entrenar el modelo, y dado que la salida que buscamos predecir se trata de una variable cualitativa, en este caso si es o no un ataque, sabemos que debemos estar ante un modelo de clasificación.

Dado que estaremos ante un modelo de clasificación binaria también revisaremos las diferentes métricas de evaluación: Accuracy, Recall y Precision. Para ello usaremos la matriz de confusión, para lograr una mejor visualización del desempeño de nuestro modelo, y la curva de ROC, que nos permitirá representar de forma visual la tasa de positivos frente a la tasa de falsos positivos. Dado que en este caso es critico identificar los ataques, priorizaremos el Recall frente a la Precision.

Aporte esperado por Big Data

Con este modelo esperamos mejorar el número de ataques identificados antes de que logren su objetivo para así mejorar la seguridad de nuestros servidores y de los datos alojados en ellos, de forma que logremos que nuestros servidores se encuentren operativos mas tiempo y de una forma más fiable, además de agilizar el bloqueo o la detección de la gran mayoría de los ataques en tiempo real, lo que además nos permitiría tomar otras medidas con estos ataques como el tratar de analizarlos en tiempo real para ver quien está detrás de dichos ataques.