PRÁCTICA MACHINE LEARNING & CIBERSEGURIDAD

CONCEPTUALICACIÓN

CASOS DE USO

NOVIEMBRE 2022

REALIZADO POR

Angie Aristizabal Bernal

KEEP CODING

2022

Descripción del caso de uso

En este informe se explicará el problema a tratar y junto a su posible solución.
Las mejoras que traerá a la empresa y lo que implicará para esta, teniendo en cuenta cómo se está afrontando ese problema actualmente
Se podrá valorar la solución para ver que tan viable es esta.

¿Cual es el problema?

Actualmente a pesar de la seguridad establecida en la empresa, quedan posibles fugas que se pueden reforzar. Como ofrecer a los clientes una mayor seguridad a sus cuentas.Teniendo en cuenta de que no hay forma de estar seguros de que cada persona haga uso de una contraseña con un gran porcentaje de seguridad y que este no esté anotada en algún papel a la vista de los demás .
El hackeo de una cuenta ya sea de un usuario o de un empleado puede comprometer la confidencialidad de los datos que se tratan y causar un gran coste monetario.
Poner en riesgo la confianza brindada por nuestros clientes al usar nuestra plataforma.
Sinse pone en juego la cuenta de un empleado se podría incluso alterar el trabajo diario de la compañía.

¿Cómo se está afrontando ahora?

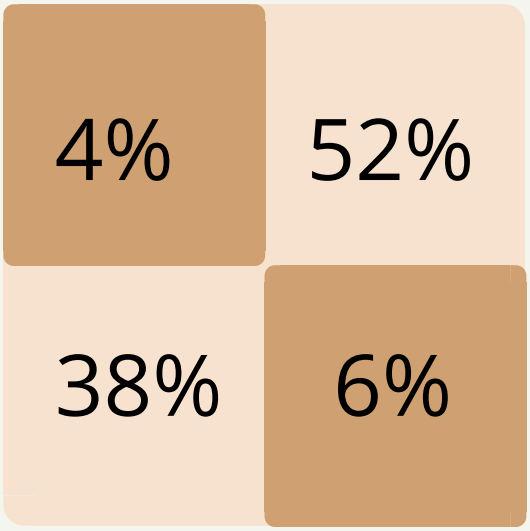
Actualmente la empresa previene el hackeo de cuentas haciendo uso de contraseñas complejas que deben cumplir los siguientes requisitos:

- Más de 8 caracteres
- El uso de mínimo una letra mayúscula
- El uso mínimos de una letra minúscula
- EL uso de caracteres especiales tales como: €,#,@,&,*,(,),=,+.

En la siguiente matriz de confusión se puede ver el porcentaje de éxito frente a las medidas tomadas hasta el momento:

Verdaderos positivos : Cuentas protegidas al 100% gracias a la complejidad desus contraseñas

Verdaderos negativos : Cuentas no protegidas por una contraseña



falsos positivos : Cuentas con 100% de seguridad por la complejidad en su contraseña. Pero una contraseña poco protegida por parte del usuario

Falsos negativos : Cuentas no protegidas por una contraseña segura. Pero si bien protegida por parte del mismo usuario

¿Que acción buscamos poder hacer para solucionar el problema ?

Incrementar la seguridad de cada una de las cuentas creando un modelo de regresión lineal que identifique las conexiones sospechosas basándose en la hora de la última conexión y en la ubicación tanto de la última conexión como de la actual. Entre más corta sea la franja de tiempo desde la última conexión a la actual y la distancia sea mayor. Más se acercará a ser clasificado como sospechoso.

KPIs - Indicadores de negocio

*Un KPI importante vendría a ser uno que nos indique que cuentas han saltado la alarma como “hackeadas”en el último mes
*Uno que nos indique cuáles de esas cuentas que han saltado la alarma, realmente fueron hackeadas
*Un indicador bastante vital sería el saber cuantas de estas cuentas hackeadas cuantas pudieron ser recuperadas gracias al aviso de nuestro modelo.
*Otro que nos indique el número de cuentas hackeadas sin haber saltado la alarma

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

Una vez implementado nuestro modelo se espera que la cifra que nos indique el número de cuentas que han saltado la alarma como hackeadas, no sea muy diferente a la cifra de cuentas que realmente han sido hackeadas.
Un mínimo esperado es la incrementacion de cuentas recuperadas gracias a la alarma de nuestro modelo y así mismo una disminución en nuestro indicador de cuentas hackeadas sin haber saltado la alarma.

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

Con el indicador de cuentas detectadas como “hackeadas” se puede permitir un margen de error del 20 %. Más el porcentaje de error aceptable al momento de recibir el indicador de cuentas que han sido realmente hackeadas es de no más que un 10%.

Experimentación: ¿Cómo vamos a corroborar el funcionamiento?

Mensualmente se deberá crear una matriz de confusión para verificar el funcionamiento de nuestro modelo . Este siempre siendo comparado con el creado previamente el mes anterior para poder comparar resultados y ver si algo ha cambiado.
Se hará uso de el valor F1 para combinar las medidas de precisión y recall en un sólo valor. Donde podremos comparar el rendimiento combinado de la precisión y la exhaustividad entre varias soluciones .

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

Se solicitará un dashboard para que de forma clara y gráfica mensualmente se reporte los resultados conseguidos con el modelo.

Equipo de trabajo

Identificación de personas colaboradoras

Se necesitará de un grupo de trabajo por parte de nuestro equipo de blue team para controlar cada alarma que surja y verificar que esta sea de una cuenta realmente hackeada. En este grupo deberá haber una persona que atienda y ayude a la recuperación de las cuentas hackeadas.
Se necesitará de otra persona que valore los riesgos de estas cuentas hackeadas y ser de soporte por sí se presentara el caso de tratarse de un cuenta de un empleado.

Detalle del caso de uso

Detalle funcional

A la hora de realizar el modelo tendrá que cuidarse el equilibrio de BIAS y varianza. Sabiendo que tratándose de un modelo de regresión lineal nuestro BIAS tendrá que estar por encima de la varianza.
Para los datos de test no se deberán usar datos de clientes recordando así la protección de datos .
Se hara uso de los diferentes rangos de IP que se pueden encontrar en distintos países.
Se le hará saber en el documento de términos y condiciones a cualquiera de nuestros usuarios que se hará uso de su dirección ip y con que fines.

Identificación de origenes de datos

Los datos para conocer la distancia entre una conexión y otra será conseguida gracias a la dirección IP de la persona conectada. Y al ya controlar el tráfico de nuestra empresa, se podrá obtener la hora de cada conexión.

Desarrollo del caso de uso

Puntos intermedios o seguimiento

Intuyo que entre mejor posición tenga el dueño de cada cuenta, mayor es su vulnerabilidad a ataques. Nuestros empleados que tienen acceso a información valiosa corren más peligro a la hora de un hackeo. Por lo que hay que concientizar a todos nuestros empleados y clientes de la importancia de asegurar nuestras cuentas lo más posible

Aporte esperado por Big Data

No se había considerado anteriormente incrementar la seguridad en este aspecto hasta la llegada de un informe del centro Nacional de ciberseguridad del Reino Unido donde se reflejaba que el 15% de los usuarios recurre a métodos seguros para proteger sus cuentas mientras que más de 40 millones de personas mantienen como contraseña de sus aparatos informáticos la más sencilla sucesión de números. A pesar de nosotros solicitar contraseñas que cumplan los ya nombrados requisitos, no podemos asegurar que sólo con eso una cuenta no sea vulnerable a hackeos. Ya que más allá de contraseñas poco seguras existen múltiples formas de recibir ataques y que se vean comprometidas nuestras cuentas.