CyberXShield: Innovant la Defensa Digital. Predicció i Mitigació de Ciberatacs Focalitzada en Patrons de Comportament

Laura Quer Calzon

Resum—Aquest projecte neix de l'observació que, malgrat els esforços actuals de les empreses per educar els seus empleats en pràctiques de seguretat de la informació, els mètodes actuals no prediuen el risc potencial de patir atacs d'enginyeria social ni consideren els patrons de comportament individuals de cada empleat. D'altra banda, hi ha metodologies d'anàlisi que podrien ser útils per a la quantificació del risc de patir atacs que utilitzen l'error humà com a principal vector d'entrada. En conseqüència, es continuen registrant nombroses pèrdues de benefici, despeses ocultes i despeses de resposta a causa de ciberatacs d'aquest tipus. En aquest treball es desenvolupa un sistema que, mitjançant l'anàlisi del comportament dels empleats en el seu entorn digital, preveu i quantifica el risc de l'organització de patir diversos ciberatacs. Addicionalment, amb aquest sistema es pretén mitigar el risc, distribuint educació granular personalitzada i generada amb intel·ligència artificial per a cada empleat. Aquesta té l'objectiu de minimitzar els comportaments de risc detectats i conscienciar als empleats de manera més efectiva.

Paraules clau—Ciberseguretat, ciberatacs, predicció, patrons de comportament, predicció, mitigació, educació granular, entorn digital, intel·ligència artificial, automatització.

Abstract— This project comes about as a result of the observation that, despite the current efforts by companies to educate their employees on information security practices, the existing methods neither predict the potential risk of being victims of social engineering attacks nor consider the individual behavior patterns of each employee. On the other hand, there are analysis methodologies that could be useful for quantifying the risk posed by cyberattacks that exploit human error as the main entry vector. Consequently, numerous losses in profit, hidden costs, and response expenses continue to occur due to such cyberattacks. This work aims to develop a system that, through the analysis of employee behavior in their digital environment, predicts and quantifies the organization's risk of being victims of certain cyberattacks. Additionally, this system aims to mitigate the risk by distributing personalized granular education generated with artificial intelligence for each employee. This aims to minimize the detected risky behaviors and raise awareness among employees more effectively.

Index Terms—Cybersecurity, cyberattacks, prediction, behavior patterns, prediction, mitigation, granular education, digital environment, artificial intelligence, automation.

1 Introducció

ACTUALMENT, garantir la seguretat digital és una de les principals preocupacions per a les organitzacions. Un ciberatac exitós pot provocar importants pèrdues financeres, danys a la reputació i repercussions legals greus, suposat conseqüències devastadores per una organització.

Al camp de la ciberseguretat, el principal vector d'entrada es troba en la propensió de les persones a cometre errors, despistar-se o no prestar atenció als detalls, una realitat que els ciberdelinqüents exploten amb freqüència [1]. Conscients que el component humà representa l'eslavó més dèbil en la defensa digital de qualsevol organització, els ciberdelinqüents usen tàctiques d'enginyeria social cada vegada més elaborades. Aquest enfocament els permet obtenir dades o portes d'entrada que posteriorment fan servir per comprometre informació confidencial, malmetre la reputació de l'empresa o extorsionar-la per obtenir un benefici econòmic.

2 Anàlisi de Requisits

2.1 Estat de l'Art

A mesura que les empreses continuen digitalitzant les seves operacions de forma exponencial, la necessitat de tractar la ciberseguretat es torna més urgent que mai. No obstant, no és suficient invertir en software i hardware segur, també és necessària una formació i capacitació dels empleats, que són la defensa més vulnerable de les organitzacions i la més explotada pels ciberdelinqüents.

Actualment, les organitzacions segueixen dues estratègies clau per quantificar i reduir el risc de patir atacs d'enginyeria social exitosos. La primera estratègia, utilitzada per quantificar el nivell de risc actual de l'organització, es basa en la simulació i monitorització d'atacs d'enginyeria social, en la majoria de casos exercicis de phishing i smishing [2].

- *E-mail de contacte: lauraquercc@gmail.com*
- *Menció realitzada: Tecnologies de la Informació*
- *Treball tutoritzat per: Hing Fai Kevin Chow (Enginyeria de la Informació i de les Comunicacions)*
- *Curs 2023/24*

La segona estratègia, tal com menciona l'Institut Nacional de Ciberseguretat (INCIBE) [3], consisteix en educar, ja sigui mitjançant cursos a aquells empleats que han sigut víctimes en les simulacions realitzades, o bé fent sessions de conscienciació genèriques per a tots els empleats de forma anual.

No obstant, tot i que les estratègies mencionades són eines molt útils per reduir el risc de patir atacs d'enginyeria social a l'entorn laboral, plantegen algunes limitacions innegables. Avui dia existeix una àmplia varietat d'atacs d'enginyeria social. Les simulacions poden abordar certs tipus d'amenaces, però no totes, deixant algunes fora d'abast pel que fa a la quantificació del nivell de risc. D'altra banda, respecte a l'estratègia d'educació i conscienciació en ciberseguretat dels empleats, es presenten diversos reptes. Les formacions molt genèriques i puntuals podrien no ser prou eficaces. A més, es dificulta la mesura de l'efectivitat d'aquestes, ja que no és possible quantificar de manera clara si han estat del tot productives.

2.2 Objectius

En aquest projecte es pretén desenvolupar un sistema de predicció i mitigació de ciberatacs que contempli i proporcioni eines per abordar les limitacions mencionades. Mitjançant un anàlisi del comportament dels empleats en el seu entorn digital, es busca predir les probabilitats que siguin víctimes de certs ciberatacs i en conseqüència mitjançant educació granular personalitzada, reduir aquests comportaments de risc que poden suposar un perjudici per a l'organització.

Amb la realització d'aquest projecte s'aspira a crear un sistema que permeti a les organitzacions satisfer els següents objectius:

- Quantificar el nivell de risc de patir diversos ciberatacs.
- Identificar activament els comportaments de risc dels empleats que posen en perill la seguretat de la informació de l'organització.
- Generar de forma personalitzada i en el moment adequat educació granular que permeti als empleats adonar-se dels comportaments de risc que tenen i puguin corregir-los.

2.3 Requisits Funcionals

Per aconseguir assolir els objectius proposats, s'han definit una sèrie de requisits funcionals que s'espera assolir al llarg del projecte.
Donat que el projecte es desenvoluparà en 4 mòduls, s'han agrupat els requisits funcionals per a cada mòdul, classificant-los segons el següent criteri:

- **Crítics (C):** Requisits essencials per al funcionament bàsic del sistema.
- **Prioritaris (P):** Requisits que agreguen valor significatiu al sistema.
- **Opcionals (O):** Requisits desitjables no fonamentals per al funcionament principal del sistema.

Mòdul 1: Identificació dels inputs.

C Determinar un conjunt de comportaments que podrien estar vinculats amb la seguretat de la informació.

Mòdul 2: Selecció i recopilació dels inputs

C Determinar quins inputs són materialment possibles de recopilar tenint en compte la tecnologia actual a l'organització.

P Donat que la pràctica d'aquest mòdul es veu restringida a causa de consideracions legals, polítiques i corporatives, simular una recopilació d'inputs.

Mòdul 3: Anàlisi i avaluació de comportaments

C Mitjançant l'anàlisi de les dades generades implementar una funcionalitat de seguiment que permeti quantificar el nivell de risc actual de l'organització.

O Generar de manera automàtica informes descriptius que proporcionin coneixement sobre el nivell de risc associat a l'organització.

Mòdul 4: Mitigació i educació personalitzada

C Generar a partir dels comportaments de risc identificats, educació granular personalitzada mitjançant eines d'intel·ligència artificial.

C Crear un procés automàtic que enviï educació granular als empleats identificats com a riscos potencials.

O Mesurar l'efectivitat de l'educació proporcionada i adaptar-la per arribar a l'objectiu de reduir el comportament de risc de l'empleat afectat.

3 Disseny

3.1 Metodologia

La metodologia seleccionada per a l'execució d'aquest projecte ha sigut la metodologia en Cascada. La raó d'aquesta elecció es basa en la clara definició i estructura del projecte, partint d'uns objectius i una planificació clara, i la necessitat d'un desenvolupament seqüencial dels mòduls proposats.

La metodologia en Cascada, adequada per projectes reduïts i ben definits, proporciona una estructura lògica i permet una progressió fluida a través de les diferents etapes del projecte [4]. Aquest enfocament es tradueix en una gestió més efectiva i eficient del projecte, maximitzant la productivitat sense comprometre el resultat final.

3.2 Planificació

Per la pròpia metodologia escollida i la possible realització d'aquest projecte, es torna essencial una correcta planifi-

cació del desenvolupament. Per aquest motiu, s'han planificat, tant les activitats pertinents de cada fase com la línia temporal, en un diagrama de Gantt.

A l'apèndix, es troba el diagrama de Gantt complet fet amb l'eina Microsoft Project (veure Fig. 1 a l'apèndix A1). Amb aquesta mateixa eina s'ha fet un seguiment del desenvolupament de les diferents fases i tasques amb l'objectiu d'assegurar una correcta progressió del projecte i l'acompliment previst de les dates d'entrega.

4 DESENVOLUPAMENT

4.1 Mòdul 1: Identificació dels inputs

Un estudi publicat a la revista HOLISTICA [5] estima que el 95% dels ciberatacs exitosos es deriven d'errors humans. En la majoria de casos, els empleats són enganyats per persones externes i acaben realitzant, sense intenció de perjudicar a l'organització, accions que posen en perill la seva seguretat. Per entendre l'arrel d'aquests comportaments de risc, s'ha recorregut a la psicologia, i a algunes de les actituds que porten a les persones a cometre'ls.

A continuació s'ha fet una correlació directa amb els trets que podrien estar vinculats amb comportaments de risc a l'hora d'interactuar en amb l'entorn digital, definint així les bases per poder identificar un conjunt d'inputs mesurables que identifiquen a una persona com a propensa a ser víctima d'un ciberatac.

- **Procrastinació:** Segons diversos estudis en psicologia [6], els individus que procrastinen, tendeixen a no seguir les polítiques de seguretat, això es tradueix en retards en actualitzacions crítiques, negligències en la gestió de contrasenyes i altres comportaments que posen en risc la seguretat de la informació.
- **Impulsivitat:** En aquesta línia Wiederhold [6] troba que l'autocontrol és una característica clau per evitar comportaments impulsius. Les persones impulsives tendeixen a clicar de forma indiscriminada en enllaços sense posar atenció a indicatius que podrien identificar una possible estafa.
- **Biaix d'optimisme:** En general, les persones tendeixen a ser optimistes i sovint subestimen la probabilitat que els passin esdeveniments indesitjables. La il·lusió d'invulnerabilitat fa que a vegades les persones comparteixin contrasenyes o ignorin advertències o recomanacions de seguretat [7].
- **Propensió al risc:** Aquest tret, caracteritza a les persones amb un excés de confiança. Es considera propensió al risc quan s'opta activament per assumir un risc, malgrat ser conscients de les conseqüències i de les accions que incrementen les probabilitats de ser víctima d'un ciberatac. Això pot conduir a decisions com no canviar les contrasenyes malgrat la possibilitat d'haver estat vulnerades, a introduir dades confidencials en llocs de poca confiança o a connectar-se a xarxes públiques tot i ser conscients dels riscos que comporta fer-ho [7].

4.2 Mòdul 2: Selecció i recopilació d'inputs

4.2.1 Selecció d'inputs

Un cop identificades les personalitats que són arrel de certs comportaments de risc molt comuns, s'ha extret el conjunt d'inputs mesurables identificats com accions o comportaments de risc. Aquests permeten quantificar, per a una organització, part del perill que suposa cada empleat.

A continuació, es llisten els inputs que s'han escollit per ser físicament possibles de recopilar.

- Edat i rol a l'organització.
- Vegades que l'empleat és víctima en un simulacre de phishing respecte les simulacions realitzades.
- Número de connexions a xarxes insegures.
- Reutilització de contrasenyes.

4.2.2 Mecanismes de recopilació d'inputs

Donat que degut a restriccions i consideracions legals, polítiques i corporatives no s'han pogut recopilar dades en en un entorn empresarial real, en aquest apartat s'explica com es recopilarien els inputs seleccionats en cas d'obtenir l'autorització pertinent. La recopilació s'ha realitzat sobre una màquina local que pretén simula l'ordinador d'un empleat.

Part dels inputs cal recopilar-los de forma manual, com el nom i cognoms de l'empleat, l'edat, el rol a l'organització, els phishings realitzats i els no superats, i el correu corporatiu. D'altra banda, els inputs que determinen patrons de comportament, com el número de connexions a xarxes insegures i els intents de reutilització de contrasenyes, cal recopilar-los de manera automàtica per poder fer una monitorització constant.

Per poder recopilar aquests darrers inputs mencionats, s'ha trobat en la monitorització dels registres d'esdeveniments de Windows un bon mètode per fer un seguiment del comportament de l'usuari. El registre d'esdeveniments de Windows fa un seguiment detallat de les activitats del sistema i proporciona informació valuosa sobre les accions realitzades per l'usuari, tornant-se ideal per a la monitorització de comportaments de risc.

Per al seguiment del número de connexions a xarxes insegures o desconegudes, el registre de Windows "WLAN-AutoConfig" [8] permet fer-ne un control. Concretament, mitjançant l'esdeveniment amb l'identificador 8001, que proporciona informació sobre totes les connexions realitzades a xarxes sense fils. A més, aquests registres especifiquen el tipus i algoritme de xifratge, així com l'autenticació que utilitzen les xarxes a les quals s'ha connectat el dispositiu, podent determinar connexions a xarxes insegures amb xifrats antics o vulnerables. A l'Apèndix A2 es pot veure un exemple de tota la informació que proporciona aquest registre, tant de la connexió a una xarxa amb un xifratge segur (veure Fig. 2), com la connexió a una xarxa sense cap mena de xifratge (veure Fig. 3).

D'altra banda, els esdeveniments "EnhancedPhishingProtection (EPP)" [9] [10] són també molt útils per a l'anàlisi del comportament, en aquest cas, la gestió de contrasenyes. Aquests events permeten registrar els intents de reutilització de contrasenyes o d'emmagatzematge local en text clar d'aquestes. Aquest esdeveniment es registra amb l'identificador 8265, especificant-se en els detalls quina ha sigut l'alerta registrada. A l'apèndix A2 (veure Fig. 4) es pot veure un exemple d'un esdeveniment causat per un intent de reutilització de contrasenya.

L'obtenció d'aquests registres d'esdeveniments s'ha programat mitjançant l'ús de llibreries i comandes integrades del sistema. En aquest cas, utilitzant la comanda integrada wevtutil [11], ha sigut possible interactuar amb el servei d'esdeveniments del sistema operatiu i consultar els registres pertinents.

4.2.3 Creació i configuració de la base de dades

Per a la gestió de les dades, s'ha optat per fer servir MySQL. En la instal·lació, s'ha utilitzat MySQL Server i juntament amb el visor SQL Workbench s'ha configurat l'entorn per a la gestió de la base de dades. Seguidament, s'ha creat una nova base de dades anomenada "cyberxshield" de tipus estructurada relacional, degut a la clara definició de l'estructura d'aquesta i la relació entre les dades que gestionarà.

Durant el desenvolupament del programa s'han creat les taules necessàries per guardar tota la informació necessària. A l'apèndix A3 (veure Fig. 5) es pot veure el diagrama entitat-relació de la base de dades final, generat amb l'eina especifica que proporciona el programa MySQL Workbench. Per al desenvolupament del mòdul 2, s'han creat les taules *employee_data* i *event_data*, que guarden respectivament totes les dades mencionades a l'apartat 4.2.1 juntament amb un identificador únic per a cada empleat i els esdeveniments registrats identificats per l'identificador de l'empleat i l'hora del registre.

Per al conjunt de proves que s'han realitzat durant el desenvolupament del sistema, s'ha registrat un únic usuari, que representa un empleat d'una empresa que implementa el sistema CyberXShield. Addicionalment, s'ha establert l'inici de la recopilació i anàlisi de registre d'esdeveniments a la data 01/12/2023 a fi de tenir un conjunt suficient per demostrar el funcionament del sistema.

4.2.4 Configuració del programa i programació de la recopilació d'inputs

L'entorn de desenvolupament integrat (IDE) utilitzat per desenvolupar el programa ha sigut PyCharm. Aquest entorn ha permès programar en Python les diferents funcionalitats requerides, com la gestió de la base de dades o la recopilació d'events de Windows entre altres, mitjançant l'ús de llibreries específiques. Tot el codi desenvolupat es pot trobar al dossier (Codi_Sistema_CyberXShield.zip).

Les classes creades en aquest mòdul s'expliquen a continuació. Addicionalment, a l'apèndix A4 (veure Fig. 6) es pot trobar el diagrama de classes complet i les relacions entre aquestes.

La classe **DbManager** actua com a administradora de la base de dades, permetent la connexió, desconnexió i manipulació de dades d'aquesta. En crear una instància DbManager amb els paràmetres requerits, que són: el nom de la base de dades, l'adreça host, que en aquest cas és el host local de la màquina, i l'usuari i la contrasenya, els quals s'han declarat com a variables d'entorn del sistema per evitar posar-les en clar al codi. El mètode *connect()* estableix la connexió. Addicionalment, en aquesta classe també s'han definit mètodes per verificar si un nou esdeveniment ja existeix a la base de dades (*isNewLog()*), per inserir nous events a la base de dades (*insertLogToBD()*) i per actualitzar la taula *employee_data* en funció de les dades recopilades i inserides a la taula *event_data* (*updateBD()*).

La classe **InputCollector** s'encarrega de recopilar les dades d'entrada, concretament el registre d'esdeveniments del sistema ja mencionats anteriorment. En crear una instància de la classe InputCollector, es proporciona un identificador d'empleat, una instància DbManager i una marca de temps que indica la data i hora a partir de la qual es vol començar a analitzar el comportament de l'empleat. La funció *collectInputs()* rep una llista de tuples especificant la llista de registres a recopilar, definint el fitxer on es troben els registres de l'esdeveniment concret i l'identificador associat a aquest. Aquesta funció actualitza la variable last_events_check amb la data actual en la què s'ha fet la última recopilació a fi d'optimitzar cada recopilació a només els registres posteriors a l'anterior recopilació. Aquesta funció també crida, per a cada registre, la funció *collectSingleInput()*, encarregada de tota la lògica d'obtenció dels events utilitzant l'execució de la comanda wevtutil com s'ha mencionat a l'apartat 4.2.2.

4.2.5 Proves unitàries mòdul 2

Per a la finalització d'aquest mòdul, a part de la comprovació del correcte funcionament de la gestió de la base de dades a través de la classe DbManager i dels mètodes de la classe InputCollector encarregats d'obtenir els registres d'esdeveniments pertinents, s'ha creat un arxiu tests.py on s'han implementat proves unitàries per validar l'apropiat resultat de les funcions *formatLog()* i *isValidLog()* de la classe InputCollector.

4.3 Mòdul 3: Anàlisi i avaluació de comportaments

Un cop configurada l'obtenció d'inputs de cada empleat, ha sigut necessari determinar la forma d'analitzar-los, el que ha portat a tractar el primer objectiu del projecte: permetre a les empreses quantificar el nivell de risc de patir diversos atacs d'enginyeria social exitosos.

4.3.1 Metodologia d'anàlisi de risc

Per quantificar de forma efectiva els riscos de seguretat de la informació associats al comportament dels empleats, s'ha enfocat l'estratègia que planteja l'Institut Nacional de Ciberseguretat, establint-se en les bases de la metodologia Magerit [12].

Aquesta metodologia es basa en el càlcul de l'impacte i la probabilitat dels riscos per quantificar el nivell de criticitat d'aquests. Però primer, ha calgut definir els escenaris de risc i les amenaces que aquests riscos comporten. Donat el

context i l'abast del projecte, així com la manca de disposició d'un entorn real, per aquest projecte, només es tindran en compte els escenaris de risc i les amenaces relacionades amb les dades recopilades al mòdul 2. Per a la identificació de les amenaces es considerarà el catàleg proporcionat a la metodologia Magerit elaborat pel Consell Superior d'Administració Electrònica i actualment mantingut per la Secretaria General d'Administració Digital amb la col·laboració del Centre Criptològic Nacional [13].

Els escenaris de risc i les amenaces associades que s'han considerat es defineixen en la Taula 1.

Escenari de Risc	Amenaces Associades
Víctima de phishing	Suplantació d'identitat, robatori de credencials, difusió de programari maliciós, frau financer
Reutilització de contrasenyes	Accés no autoritzat, fuga d'informació
Connexió a xarxes insegures	Intercepció d'informació, exposició a programari maliciós

Taula 1: Escenaris de risc i amenaces associades

Un cop identificades les amenaces que comporten els escenaris de risc plantejats, s'ha avaluat el seu risc. Per a cada escenari de risc, s'ha estimat la probabilitat que es materialitzi i l'impacte que tindria sobre el negoci si l'amenaça fos exitosa. La Taula 2 i la Taula 3 permeten classificar els escenaris de risc tan qualitativament com quantitativament.

Qualitatiu	Quantitatiu	Descripció
Baix	1	El risc es materialitza cada 365 dies o més.
Mig	2	El risc es materialitza cada 30 dies o més.
Alt	3	El risc es materialitza cada 7 dies o més.
Crític	4	El risc es materialitza cada 6 dies o menys.

Taula 2: Càlcul de la probabilitat

Qualitatiu	Quantitatiu	Descripció
Baix	1	El dany associat a la materialització de l'amenaça no te conseqüències rellevants per a l'organització.
Mig	2	El dany associat a la materialització de l'amenaça te conseqüències notables per a l'organització.
Alt	3	El dany associat a la materialització de l'amenaça te conseqüències greus per a l'organització
Crític	4	El dany associat a la materialització de l'amenaça te conseqüències desastroses per a l'organització.

Taula 3: Càlcul de l'impacte

A continuació, a la Taula 4 s'han definit els impactes associats a cada escenari-amenaça per a l'exemple que s'ha tractat en aquest projecte.

Escenari de Risc	Amenaces Associades	Impacte per amenaça	Impacte per escenari
Víctima de phishing	Suplantació d'identitat	2	4
	Robatori de credencials	4	
	Difusió de programari maliciós	3	
	Frau financer	4	
Reutilització de contrasenyes	Accés no autoritzat	4	4
	Fuga d'informació	3	
Connexió a xarxes insegures	Intercepció d'informació	2	3
	Exposició a programari maliciós	3	

Taula 4: Càlcul de l'impacte

Un cop identificat l'impacte que suposa cada risc, s'ha calculat la probabilitat que els riscos succeïssin. El valor de la probabilitat de la reutilització de contrasenyes i la connexió a xarxes insegures s'ha determinat seguint la Taula 2, però per al risc de víctima de phishing s'ha fet una lleugera modificació, ja que depèn dels exercicis de phishing que ha realitzat l'empresa. Aquesta es pot trobar a la Taula 5.

Qualitatiu	Quantitatiu	Descripció
Baix	1	El risc s'ha materialitzat un sol cop.
Mig	2	El risc s'ha materialitzat menys de la meitat del total d'exercicis.
Alt	3	El risc s'ha materialitzat més de la meitat del total d'exercicis.
Crític	4	El risc s'ha materialitzat tots els cops que s'ha realitzat l'exercici.

Taula 5: Càlcul de probabilitat per a riscos que depenen de la realització d'exercicis

Per últim, per al càlcul de risc total, s'ha utilitzat la formula risc = probabilitat + impacte, sent la probabilitat la mitjana de tots els empleats.

La Taula 6 determina el nivell de risc segons el valor obtingut. On s'ha fet l'associació de risc següent: 2/3 – Molt Baix, 4 – Baix, 5 – Mig, 6 – Alt, 7/8 – Crític.

Càlcul de risc					
Probabilitat	4	5	6	7	8
	3	4	5	6	7
	2	3	4	5	6
	1	2	3	4	5
		1	2	3	4
		Impacte			

Taula 6: Interpretació del resultat del risc

4.3.2 Programació de l'automatització de l'anàlisi

La classe **DataAnalyzer** (veure Fig. 6 a l'apèndix A4 per més detalls) s'ha desenvolupat per fer l'anàlisi automatitzat de les dades recopilades i obtenir els resultats de risc seguint la metodologia explicada a l'apartat 4.3.1.

El constructor d'aquesta classe rep la instancia DbManager amb la que s'ha realitzat la connexió i una llista de tuples que contenen el següent: nom del risc, identificador de l'esdeveniment o columna de la base de dades on es troba el total d'exercicis realitzats, i nº de taula per al càlcul de probabilitat. Els dos mètodes principals d'aquesta classe són *individualAlalysis()*, que calcula per a cada empleat la probabilitat de risc i ho actualitza a la taula *analysis_data* de la base de dades; i el mètode *calculateRisk()*, al que per paràmetre se li passa una llista de tuples amb el risc i l'impacte de cada un i actualitza les taules *global_risk_data* i *matrix_data* a la base de dades.

La taula *global_risk_data* guarda un resum dels resultats dels riscos monitorats i obtinguts. D'altra banda, la taula *matrix_data* emmagatzema per a cada risc, l'impacte i la probabilitat global obtinguda. Per poder diferenciar les dades recopilades a cada report realitzat, aquestes s'agrupen utilitzant la data i hora en la qual s'ha realitzat l'anàlisi.

4.3.3 Visualització interactiva de l'informe en Power BI

L'últim pas d'aquest mòdul, i amb el qual s'ha aconseguit complir el segon objectiu del projecte: identificar i visualitzar activament els riscos que posen en perill la seguretat de la informació de l'organització, s'ha realitzat amb l'eina Power BI [14]. Aquest programa és una eina nativa de Microsoft que, entre altres, permet convertir una base de dades en un dashboard que mostra informació coherent, interactiva i atractiva visualment.

El dashboard que s'ha creat per poder visualitzar un informe interactiu de les dades analitzades es troba a l'apèndix A5 (veure Fig. 7). A més també està inclòs l'arxiu Risk_Assessment_CyberXShield_(Dashboard_PowerBI)-.pbix adjunt en el dossier. En el moment de la generació de l'arxiu, s'han realitzat 3 informes de forma manual per tal de veure com es visualitzaria l'històric després de tres mesos.

El dashboard desenvolupat consta de 4 parts:

- **Valor per escenari de risc:** que conté una gràfica on es llisten tots els riscos monitoritzats amb el seu valor de risc calculat. El color de cada barra varia segons el valor de risc.
- **Resum de dades generals:** format per 4 targetes d'informació, es mostren dades rellevants. Concretament, el risc global, el número de riscos controlats, el número de riscos crítics i el número d'empleats monitorats.
- **Matriu de calor per a la categorització de riscos:** la matriu implementada, no només mostra l'impacte i probabilitat de cada risc, sinó que també defineix la interpretació de criticitat d'aquests, especificant els nivells des de molt baix a crític. A més, el nivell qualitatiu es mostra en passar amb el ratolí per sobre de cada quadrant de la matriu.
- **Tendència de risc:** aquest darrer gràfic lineal mostra un històric del valor de risc dels riscos monitorats, permetent portar un control més precís i detallat de l'evolució dels riscos al llarg del temps.

Addicionalment, el dashboard compta amb diversos punts d'interacció:

- **Selecció de la data de l'informe:** aquest desplegable situat a la cantonada superior dreta, permet seleccionar la data en la qual es vol mostrar les dades. En canviar-la, s'actualitzen automàticament el dashboard.
- **Gràfica de valor per escenari de risc:** en seleccionar la barra d'algun dels riscos de la gràfica, de forma automàtica s'actualitza la matriu de calor per mostrar l'impacte i la probabilitat del risc concret.
- **Mode d'enfocament:** totes les gràfiques consten amb l'opció de visualitzar en mode d'enfocament, el qual desplega la gràfica seleccionada en un format ampliat, que permet una inspecció més detallada dels elements.
- **Crear alertes:** és possible crear alertes per a qualsevol element del dashboard, permetent definir condicions específiques i rebre notificacions immediates en cas que es compleixin.

4.3.4 Proves unitàries mòdul 3

En aquest mòdul, la principal funcionalitat que s'ha desenvolupat ha sigut el càlcul de risc a la classe DataAnalyzer i la generació de l'informe dinàmic en PowerBI directament connectat a la base de dades. Donat que els mètodes de la classe DataAnalyzer han constat bàsicament de funcions dedicades a realitzar els càlculs explicats a l'apartat 4.3.1 i guardar els valors a la base de dades, no s'han considerat necessàries proves unitàries. A més, donat que totes les funcions de la classe estan vinculades estretament a la base de dades, elaborar proves unitàries efectives amb Mocks per simular les interaccions amb aquest, s'ha considerat que no proporcionava un valor significatiu i, per tant, s'ha obviat.

4.4 Mòdul 4: Mitigació i educació personalitzada

Aquest sistema no només pretén proporciona a les empreses la capacitat de quantificar el nivell de risc actual al qual estan exposades pel comportament dels seus empleats, sinó que també busca intervenir activament per mitigar aquest risc.

Amb totes les dades sobre el risc actual de cada empleat a l'abast, emergeix de manera inherent la capacitat de prendre accions per reduir aquest risc. D'aquesta manera, en aquest últim mòdul es desenvoluparà el procediment per fer això realitat, complit així l'últim objectiu del projecte: generar de forma personalitzada i en el moment adequat educació granular que permeti als empleats adonar-se dels comportaments de risc que tenen i puguin corregir-los.

4.4.1 Enviament automàtic de correus electrònics

En aquest punt, cal destacar que, a causa de les limitacions actuals del projecte i el fet que no es desenvolupa en un entorn empresarial real, s'ha optat per la utilització d'un compte de Gmail en lloc de comprar, configurar i utilitzar un domini propi per a l'enviament de l'educació granular a través de correu electrònic. El correu de Gmail creat, configurat i utilitzat per a aquest projecte ha sigut cyberXShieldAwareness@gmail.com.

D'aquesta forma, l'enviament de correus, a través de la adreça de correu mencionada, s'ha realitzat mitjançant la llibreria stmplib de Python [15] [16]. Aquesta llibreria no només permet enviar correus formatats amb HTML de manera automatitzada, sinó que a més especifica el protocol de seguretat a utilitzar, en aquest cas TLS. El codi que manega la funcionalitat d'enviament automàtic, es pot trobar a la classe **AwarenessManager** (veure Fig. 6 a l'apèndix A4 per més detalls), concretament al mètode *sendPill()*.

4.4.2 Disseny de l'educació granular

Abans de generar el contingut dinàmic de l'educació granular que s'enviarà per correu als empleats als què se'ls hagi identificat un comportament de risc, s'ha desenvolupat en HTML una plantilla amb l'estil formatat i els camps pertinents que variaran segons el risc detectat.

Concretament, el correu conté un text curt explicatiu, on es menciona que s'ha detectat un comportament de risc i que a continuació es mostra una guia de per què suposa un risc i que fer per corregir-lo. Seguidament, es mostra el risc detectat i a continuació tres apartats amb varies frases fent referència als perills d'aquest comportament, que fer per corregir-lo i accions a evitar relacionades. A l'apèndix A6 (veure Fig. 8 i Fig. 9) es pot veure amb més detall el format generat.

4.4.3 Integració d'intel·ligència artificial al sistema

La integració de la intel·ligència artificial al sistema CyberXShield és el darrer pas que conclou aquest projecte. Aquesta incorporació permet que les píndoles educatives, enviades pel sistema quan es detecten comportaments de risc, siguin generades de manera personalitzada i dinàmica. Això, no només millora la precisió i eficàcia del sistema per mitigar els comportaments de risc, sinó que també permet una evolució constant en paral·lel, amb els avenços de la intel·ligència artificial.

Per a aquest projecte, s'ha utilitzat Hugging Face Transformers [17], un marc de codi obert dedicat a construir i compartir models d'aprenentatge automàtic. La plataforma conté una àmplia varietat de models preentrenats, a més, permet accedir a aquests a través d'una API. Això ha fet possible que, tot i tenir restriccions de pressupost i de còmput, hagi sigut possible introduir la intel·ligència artificial en aquest projecte com a mecanisme per generar de forma automàtica contingut educatiu dinàmic i adaptat a cada empleat.

El primer pas d'aquesta integració ha sigut l'elecció del model més adequat per a la tasca requerida. En aquest cas, la de generar de forma automàtica un conjunt de frases sobre els perills, les bones pràctiques i accions a evitar, respecte a un comportament de risc, adaptant aquestes frases a l'edat i el rol de l'empleat. Dins l'àmbit del processament del llenguatge natural, Hugging Face proporciona una classificació de models molt àmplia. Dins d'aquesta categorització, els models de generació de text són els més rellevants per aquest projecte, donat que poden produir text coherent i rellevant a partir d'un context donat. Tot i això, per a la tasca plantejada, la categoria més adequada és la de generació de text basada en instruccions. Aquests models estan dissenyats per produir contingut personalitzat i contextualitzat seguint pautes específiques [18].

Un cop definida la categoria del model requerida, s'ha definit un prompt per poder testejar diferents models i concloure amb quin s'obté el millor resultat. El prompt proporcionat als models ha sigut el següent:

"An employee with following characteristics: age [edat de l'empleat], position: [rol de l'empleat], has been detected with risky behavior of [risc detectat]. Provide exactly 4 distinct and very short sentences for an educational pill explaining why this behavior poses a risk to the company. Then generate other 4 distinct and very short sentences listing how the employee could correct this risk behaviour. To finalize, generate other 4 distinct and very short sentences listing which behaviour should avoid the employee related with the risk behaviour detected. To be able to interpret the output returns a Python dictionary with the keys 'risks', 'correct', 'avoid', where the values are respectively lists of strings with the phrases of each category. Additionally, inside the string of every sentence, use the tags and to bold rellevant words of the sentence.".

És important destacar que, per interpretar correctament el resultat retornat pel model d'intel·ligència artificial, s'ha sol·licitat un diccionari en el llenguatge Python.

Després de realitzar diverses proves a través de la plataforma Hugging Face amb diferents models de generació de text basada en instruccions, s'ha trobat que el model mistralai/Mistral-7B-Instruct-v0.2 [19] era el que generava la millor sortida. Donat que era el que millor seguia les instruccions de com havia de ser el format de sortida per ser interpretat correctament ha sigut el model final escollit.

Per a la integració del model seleccionat i la interacció amb aquest, a la classe AwarenessManager, s'han creat les funcions següents:

- **createPrompt(risk, age, role):** modifica el prompt segons les característiques de l'empleat i el risc detectat.
- **generateAiText(prompt):** s'especifica el token_access, necessari per accedir a l'API de Hugging Face, i el model a utilitzar. Finalment, mitjançant una consulta POST, retorna el text generat pel model.
- **interpretOutput(ai_output):** obté el text generat pel model i l'interpreta per organitzar les frases generades. Crea un diccionari on cada valor és una llista de les frases corresponents a cada apartat de la píndola educativa.
- **assembleOutput(result_dict, risk):** Per últim, aquesta funció concatena el cos de la plantilla HTML de la píndola amb les frases generades per la intel·ligència artificial.

Addicionalment, cal mencionar que per obtenir un bon resultat del model, a part del prompt també ha sigut necessari especificar les propietats de l'output generat. Concretament, s'han especificat els següents paràmetres:

- **max_new_tokens:** 600, que ampliar el número de tokens de sortida generats pel model, evitant que la resposta no es completi.
- **temperature:** 0.7, que introdueix un nivell moderat d'aleatorietat al text per crear respostes més diverses i creatives.
- **top_p:** 0.9, que controla la fluïdesa i coherència del text, seleccionant les paraules més probables.
- **top_k:** 500, que restringeix la selecció de paraules a les 500 més probables per garantir la rellevància semàntica i la cohesió del text final.

A l'apèndix A6 (veure Fig. 8 i Fig. 9) es poden trobar dos exemples del resultat final de l'educació granular enviada on la informació ha sigut generada per intel·ligència artificial

4.4.4 Proves unitàries mòdul 4

Per aquest últim mòdul les proves unitàries que s'han desenvolupat s'han centrat en la nova classe AwarenessManager i que es poden trobar dins el fitxer tests.py. Els testos desenvolupats són les següents:

- **Test_employeesToSend:** Aquest test verifica que el format retornat per la funció *employeesToSend()* és correcte. Concretament, una llista de diccionaris amb la informació corresponent de l'empleat.
- **Test_sendPill:** Aquest altre test comprova que la funció *sendPill(),* donat un receptor, envia correctament l'educació granular i, per tant, retorna True. En aquest punt s'ha detectat que el model d'intel·ligència artificial no sempre retornava el format esperat i, que per tant, no era interpretable pel codi. Per això s'han realitzat correccions a la funció que comproven si el model retorna el format esperat. En cas de no ser així, es torna a fer la petició fins a obtenir un resultat interpretable amb un màxim de 10 intents.

Cal mencionar que amb models més avançats d'intel·ligència artificial les probabilitats d'obtenir formats no vàlids i diferents de l'especificat al prompt redueixen, millorant així l'eficàcia del sistema.

Addicionalment, en realitzar les peticions repetidament quan fallava el format, també s'ha detectat que la resposta obtinguda pel model d'intel·ligència artificial era exactament la mateixa en tots els casos, per això, s'ha incorporat, al mètode *generateAiText(),* una línia de codi que afegia un número aleatori entre 0 i 1000 al final del prompt passat per paràmetre. Això ha permès que el model interpretés el prompt com a nou i que, per tant, generes una nova resposta diferent.

5 PROVES DE SISTEMA I RESULTATS

En el context d'aquest projecte, per finalitzar el desenvolupament i demostrar els resultats, s'han realitzat proves de sistema per validar el funcionament integral de tots els mòduls desenvolupats, en condicions simulades que reflecteixen el seu ús en un entorn més real.

Durant el procés de desenvolupament, s'ha treballat amb un únic usuari a fi de focalitzar l'atenció en les funcionalitats del sistema, però un cop integrats tots els mòduls i validat el correcte funcionament d'aquests, així com l'assoliment dels objectius proposats, s'han procedit a realitzar les proves de sistema pertinents per verificar una correcta escalabilitat a un major número d'usuaris.

Per realitzar les proves amb múltiples usuaris, s'ha configurat una instància virtual de Windows utilitzant VirtualBox, què ha simulat la presència d'un segon usuari al sistema. Per executar el programa des de la nova màquina, ha calgut crear un nou usuari Mysql amb accés a la base de i permís per connectar-se només des de la ip de la màquina virtual. També se li han donat privilegis sobre la base de dades i s'ha modificat el fitxer my.ini per permetre connexions remotes.

Addicionalment, s'ha afegit el nou empleat a la base de dades amb les següents dades rellevants: identificador: 2, rol: president and CEO, edat: 59, phishings fallats: 3 i phishings totals realitzats: 5. La incorporació d'aquest nou usuari ha permès verificar com la recopilació d'inputs, la generació d'informes i l'enviament de píndoles personalitzades s'ha realitzat correctament. Adjunt al dossier es pot trobar l'informe en PowerBI Risk_Assessment_CyberXShield_(Dashboard_PowerBI).pbix on seleccionant la data més recent es pot observar com s'actualitzen les dades després d'afegir un nou usuari.

A continuació es mostra l'eficàcia dels resultats obtinguts en les píndoles generades, a través de dos casos del mateix risc detectat, concretament el de phishings no superats, per

als dos usuaris monitorats. Es demostra com la utilització d'un model d'intel·ligència artificial adapta l'educació granular segons les característiques de l'empleat. Seguidament, es mostra com s'ha generat de forma personalitzada, la informació més rellevant per a l'empleat objectiu, permetent, per tant, una major eficàcia del mètode de mitigació.

Per al primer cas, l'objectiu de la píndola ha sigut un empleat de 21 anys amb el rol "intern in the sales department at Marsh" i s'ha obtingut el següent resultat (veure Fig. 10)

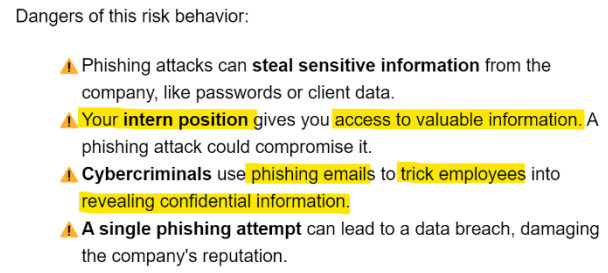


Fig 10. Resultat de la píndola per a l'empleat amb id 1

El segon cas ha consistit en un empleat de 59 anys amb el rol "president and CEO at Marsh" i s'ha obtingut el següent resultat (veure Fig. 11).

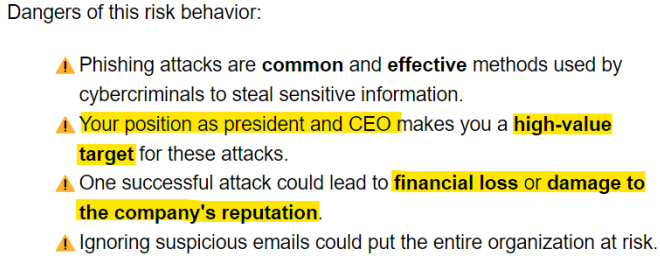


Fig 11. Resultat de la píndola per a l'empleat amb id 2

Les diferències en les frases generades en ambdós casos, demostren personalització segons l'edat i el rol. Per al becari, el to és més directe i educatiu, emfatitzant la vulnerabilitat de la seva posició i la importància de protegir la informació a la qual té accés. En contrast, per al president i CEO, el to és més seriós i formal, subratllant l'alta visibilitat del rol, els riscos financers i de reputació, i la responsabilitat de protegir tota l'organització, ressaltant la gravetat d'ignorar correus sospitosos.

Durant el projecte s'han culminat tots els objectius plantejats. S'han aconseguit identificar automàticament comportaments de risc plasmant els resultats en un informe interactiu i dinàmic, permetent quantificar el nivell de risc global a l'organització. A més, s'ha desenvolupat la generació automatitzada d'educació granular de manera personalitzada i en el moment adequat gràcies a eines d'intel·ligència artificial. Això, ha permès crear un sistema de mitigació que aspira a ser significativament més efectiu que els mètodes actualment utilitzats.

Respecte als requisits funcionals plantejats a l'inici del projecte, s'han complert tant els crítics com els prioritaris tal com s'esperava a la planificació. Al mòdul 1, s'ha realitzat una investigació exhaustiva per identificar quins comportaments estan vinculats amb la ciberseguretat, determinant que la procrastinació, la impulsivitat, el biaix d'optimisme i la propensió al risc són els més rellevants. Al mòdul 2, a part de determinar quins inputs són materialment possibles de recopilar, s'ha aconseguit programar una recopilació automàtica dels inputs que identifiquen els comportaments de risc plantejats en aquest projecte. Al mòdul 3, seguint el mètode Magerit d'anàlisi de riscos, s'ha implementat una funcionalitat de seguiment que ha permès quantificar el nivell de risc actual a l'organització. A més, s'ha complert el requisit opcional de generar un informe dinàmic i descriptiu que permetés visualitzar les dades recopilades. Aquest informe, creat en Power BI, s'actualitza de forma automàtica i permet identificar més ràpidament riscos potencials en el comportament dels empleats. Finalment, al mòdul 4, s'ha generat educació granular personalitzada mitjançant models d'intel·ligència artificial proporcionats per Hugging Face, i s'ha creat un procés automàtic d'enviament d'educació granular als empleats identificats com a riscos potencials. El requisit opcional d'aquest mòdul, que consisteix a mesurar l'efectivitat de l'educació granular proporcionada i adaptar-la per reduir el comportament de risc en els empleats afectats, s'ha posposat per a etapes futures del desenvolupament del projecte. Quan es disposi d'un conjunt d'usuaris reals per a proves i d'un model d'intel·ligència artificial més avançat, els resultats d'aquest requisit seran més rellevants, permetent mesurar de manera efectiva la capacitat de mitigar els comportaments de risc dels empleats afectats.

6 Conclusió i línies futures

Aquest projecte ha demostrat posar èmfasis en la importància del tractament del principal vector d'entrada en la gran majoria dels ciberatacs a les organitzacions: el factor humà i els seus comportaments de risc. S'ha destacat com, els mètodes convencionals no consideren les particularitats individuals dels empleats, i s'ha proposat un mètode innovador de predicció i mitigació de ciberatacs focalitzat en patrons de comportament.

Com a possibles millores del projecte i línies de continuació, es proposa l'ús d'una base de dades al núvol, permetent un accés més ràpid i eficient a la informació, així com una major escalabilitat en un entorn real on la quantitat d'empleats monitorats ascendeix significativament. Addicionalment, monitorar un major nombre de riscos podria proporcionar una visió més completa i detallada del risc actual a l'organització.

Degut a restriccions de pressupost, no s'han pogut utilitzar els models més potents d'intel·ligència artificial. Tot i això, els resultats obtinguts amb els recursos disponibles han estat molt prometedors. Aquesta limitació obre una possible àrea de millora, on l'ús de models més avançats podria potenciar encara més l'eficàcia del sistema en la generació d'educació granular personalitzada. Aquests models no només millorarien la qualitat de les píndoles educatives enviades als empleats, sinó que també permetrien analitzar totes les píndoles enviades anteriorment i seguir-ne

l'efectivitat. Això contribuiria a generar píndoles encara més efectives en el futur, optimitzant contínuament el procés de mitigació de riscos.

Aquest projecte subratlla la necessitat d'una estratègia adaptada als individus per enfortir la protecció digital en un entorn corporatiu cada cop més complex i dinàmic, i amb els resultats obtinguts demostra que és possible una implementació real d'aquesta solució.

AGRAÏMENTS

M'agradaria expressar el meu agraïment a totes les persones que han contribuït de manera significativa a la realització d'aquest projecte.

En primer lloc, agraeixo al meu tutor, Kevin Chow, per confiar en mi i oferir-me l'oportunitat de dur a terme aquest projecte. També vull expressar la meva gratitud a les meves mentores a l'empresa Marsh, la Nelia Argaz i la Andrea López, per la seva assistència tant en la concepció de la proposta, com en el desenvolupament del projecte.

Finalment, dedico un especial agraïment als meus amics i familiars pel seu constant suport i confiança que m'han permès continuar donant el màxim per arribar als millors resultats.

BIBLIOGRAFIA

[1] D. G. "Error Humano: La razón principal de las Brechas de Seguridad", Interbel, [En línia] Disponible a: https://www.interbel.es/error-humano/

[2] "Ejemplos De Ingeniería Social En El Ámbito Laboral: Cómo Prevenir Los Ataques Y Proteger La Información De Tu Empresa", Vanbig, [En línia] Disponible a: https://vanbig.es/ciberseguridad/ejemplos-de-ingenieria-social-en-el-ambito-laboral-como-prevenir-los-ataques-y-proteger-la-informacion-de-tu-empresa/

[3] "TemáTICas Ingeniería social", INCIBE, [En línia] Disponible a: https://www.incibe.es/empresas/tematicas/ingenieria-social#concienciacion

[4] Y. Bassil, "A Simulation Model for the Waterfall Software Development Life Cycle", (iJET), Vol. 2, No. 5, Maig 2012 [En línia] Disponible a: https://arxiv.org/abs/1205.6904

[5] C. Nobles, "Botching Human Factors in Cybersecurity in Business Organizations", (HOLISTICA), Vol. 9, No. 3, pp. 71-88, Desembre 2018, [En línia] Disponible a: https://doi.org/10.2478/hjbpa-2018-0024

[6] A. Moustafa, A. Bello, A. Maurushat, "The Role of User Behaviour in Improving Cyber Security Management", (Frontiers in Psychology), Vol. 12, Juny 2021, [En línia] Disponible a: https://doi.org/10.3389/fpsyg.2021.561011

[7] O. Bravo, "Ciberseguridad para escépticos", [En línia] Disponible a: https://www.debatesiesa.com/ciberseguridad-para-escepticos/

[8] What2Log, "Wifi Connection", [En línia] Disponible a: https://what2log.com/windows/logs/win10wificonnection/

[9] S. Hamilton, "Protect passwords with enhanced phishing protection", [En línia] Disponible a: https://techcommunity.microsoft.com/t5/windows-it-pro-blog/protect-passwords-with-enhanced-phishing-protection/ba-p/3631881

[10] Windows, "Enhanced Phishing Protection in Microsoft Defender SmartScreen", [En línia] Disponible a: https://learn.microsoft.com/en-us/windows/security/operating-system-security/virus-and-threat-protection/microsoft-defender-smartscreen/enhanced-phishing-protection?tabs=intune

[11] Ss64, "WEVTUTIL", [En línia] Disponible a: https://ss64.com/nt/wevtutil.html

[12] INCIBE, "Análisis de riesgos en 6 pasos", [En línia] Disponible a: https://www.in-cibe.es/empresas/blog/analisis-riesgos-pasos-sencillo

[13] Secretaría General de Administración Digital , "MAGERIT v.3 : Metodología de Análisis y Gestión de Riesgos de los Sistemas de Información", [En línia] Disponible a: https://ad-ministracionelectronica.gob.es/pae_Home/pae_Documentacion/pae_Metodo-log/pae_Magerit.html

[14] D. Iseminger , "¿Qué es Power BI? | Microsoft Learn", [En línia] Disponible a: https://le-arn.microsoft.com/es-es/power-bi/fundamentals/power-bi-overview

[15] "smtplib — SMTP protocol client — Python 3.12.3 documentation", [En línia] Disponi-ble a: https://docs.python.org/3/library/smtplib.html

[16] O. Mishra, "How to Send Beautiful Emails in Python" [En línia] Disponible a: https://www.geeksforgeeks.org/how-to-send-beautiful-emails-in-python/

[17] Mssaperla, "¿Qué son los Hugging Face Transformers? | Microsoft Learn", [En línia] Dis-ponible a: https://learn.microsoft.com/es-es/azure/databricks/machine-learning/train-model/huggingface/

[18] Hugging Face, "What is Text Generation?", [En línia] Disponible a: https://hugging-face.co/tasks/text-generation

[19] Mistral AI Team, "mistralai/Mistral-7B-Instruct-v0.2 | Hugging Face", [En línia]. Disponi-ble a: https://hugging-face.co/mistralai/Mistral-7B-Instruct-v0.2

APÈNDIX

A1. DIAGRAMA DE GANTT

Id	Nom de tasca	Duració	Inici	Fi
1	**Sistema CyberXShield**	**145 días**	**lun 12/2/24**	**vie 5/7/24**
2	**Fase 1: Anàlisi de requisits**	**12 días**	**lun 12/2/24**	**vie 23/2/24**
3	Investigació de mercat	7 días	lun 12/2/24	dom 18/2/24
4	Definició i classificació d'objectius i requisits	5 días	lun 19/2/24	vie 23/2/24
5	**Fase 2: Disseny i planificació**	**10 días**	**sáb 24/2/24**	**lun 4/3/24**
6	Disseny detallat del sistema	10 días	sáb 24/2/24	lun 4/3/24
7	Elecció de la metodologia	4 días	sáb 24/2/24	mar 27/2/24
8	Planificació	6 días	mié 28/2/24	lun 4/3/24
9	**Fase 3: Desenvolupament i proves unitàries**	**98 días**	**mar 5/3/24**	**lun 10/6/24**
10	Mòdul 1: Identificació dels inputs	8 días	mar 5/3/24	mar 12/3/24
11	**Mòdul 2: Selecció i recopilació dels inputs**	**24 días**	**mié 13/3/24**	**vie 5/4/24**
12	Creació i configuració de la BD	8 días	mié 13/3/24	mié 20/3/24
13	Creació d'usuaris test i columnes input	3 días	jue 21/3/24	sáb 23/3/24
14	Programació de script de generació d'inputs	8 días	dom 24/3/24	dom 31/3/24
15	Proves unitàries del mòdul 2	5 días	lun 1/4/24	vie 5/4/24
16	**Mòdul 3: Anàlisi i avaluació de comportaments**	**27 días**	**sáb 6/4/24**	**jue 2/5/24**
17	Creació i configuració del programa	3 días	sáb 6/4/24	lun 8/4/24
18	Funcionalitat detecció i seguiment de comportaments de risc	7 días	mar 9/4/24	lun 15/4/24
19	Agregació de taula d'anàlisis a la BD	3 días	mar 16/4/24	jue 18/4/24
20	Funcionalitat guardar dades a la BD	6 días	vie 19/4/24	mié 24/4/24
21	Funcionalitat crear informe de probabilitats	8 días	jue 25/4/24	jue 2/5/24
22	~~Proves unitàries i d'integració del mòdul 3~~	~~5 días~~	~~vie 3/5/24~~	~~mar 7/5/24~~
23	**Mòdul 4: Mitigació i educació personalitzada**	**39 días**	**vie 3/5/24**	**lun 10/6/24**
24	Integració i comprenció de l'API GPT	12 días	vie 3/5/24	mar 14/5/24
25	Funcionalitat de generació de píldores	12 días	mié 15/5/24	dom 26/5/24
26	Implementació SMTP i SSL/TLS per l'enviament de píldores	10 días	lun 27/5/24	mié 5/6/24
27	Proves unitàries i d'integració del mòdul 4	5 días	jue 6/6/24	lun 10/6/24
28	Fase 4: Proves de sistema	7 días	mar 11/6/24	lun 17/6/24
29	**Fase 5: Tancament**	**18 días**	**mar 18/6/24**	**vie 5/7/24**
30	Revisió final de la documenació	8 días	mar 18/6/24	mar 25/6/24
31	Preparació de la defensa	10 días	mié 26/6/24	vie 5/7/24

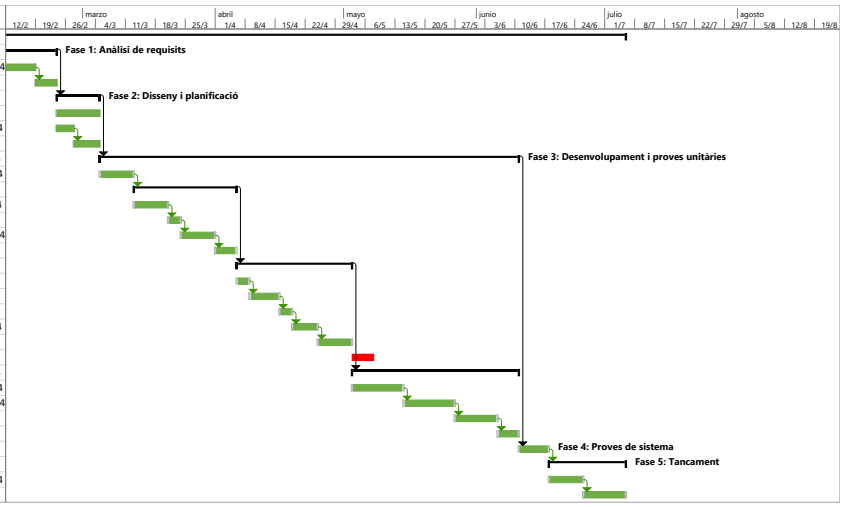


Fig 1. Diagrama de Gantt

A2. REGISTRES D'ESDEVENIMENTS DE WINDOWS

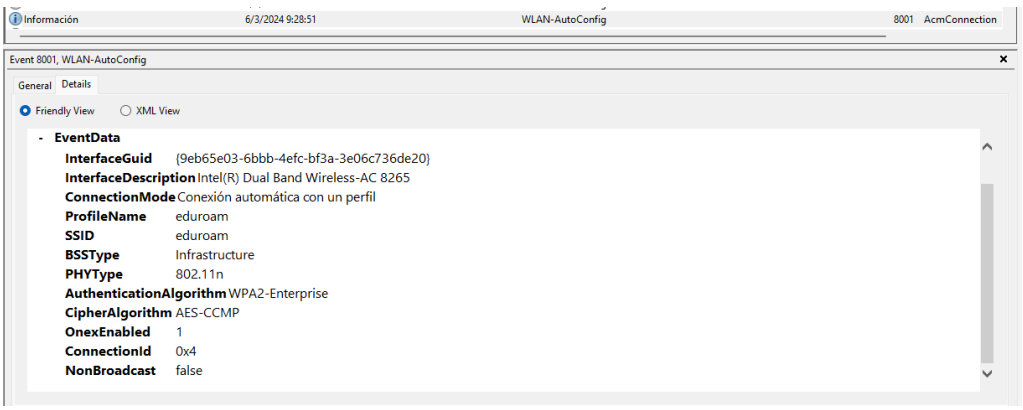


Fig 2. Detalls del registre 8001(WLAN-AutoConfig) d'un esdeveniment amb una connexió segura

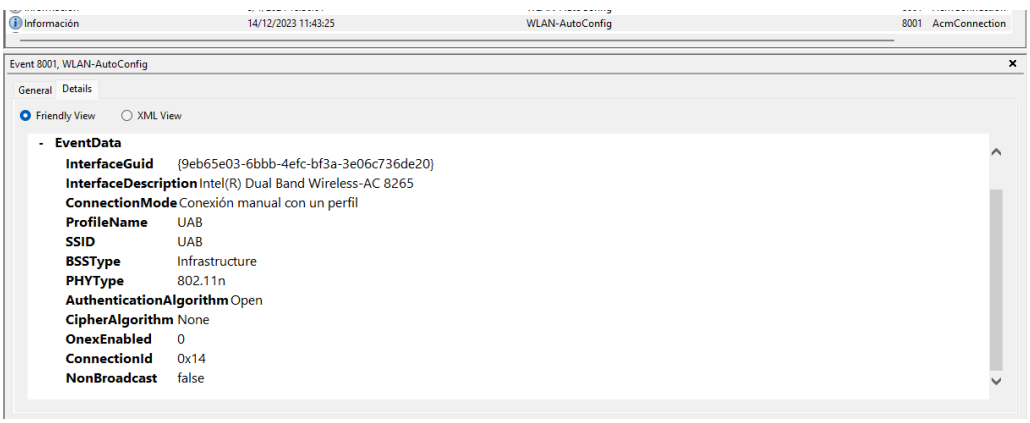


Fig 3. Detalls del registre 8001(WLAN-AutoConfig) d'un esdeveniment amb una connexió insegura

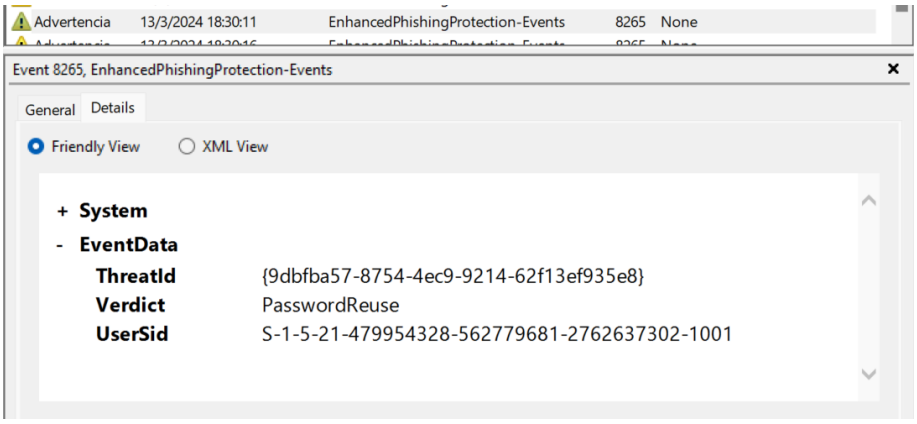


Fig 4. Detalls del registre 8265 (EnhancedPhishingProtection) d'un esdeveniment d'intent de reutilització de contrasenya

A3. DIAGRAMA ENTITAT-RELACIÓ DE LA BASE DE DADES

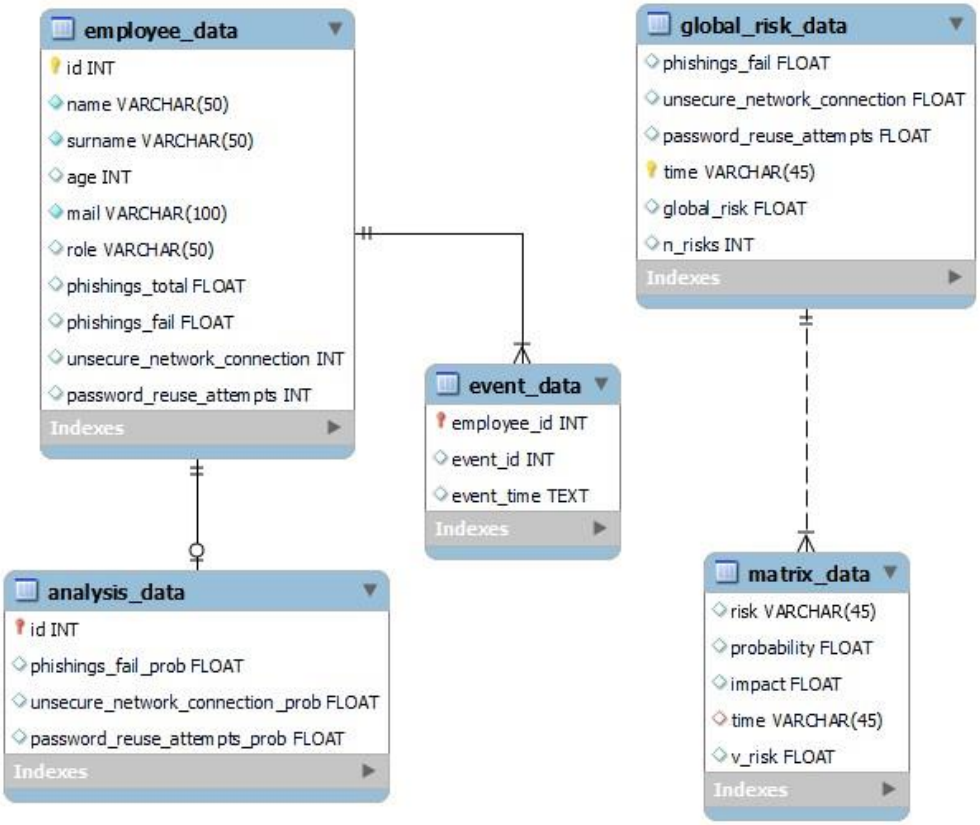


Fig 5. Diagrama entitat-relació de la base de dades

A4. DIAGRAMA DE CLASSES

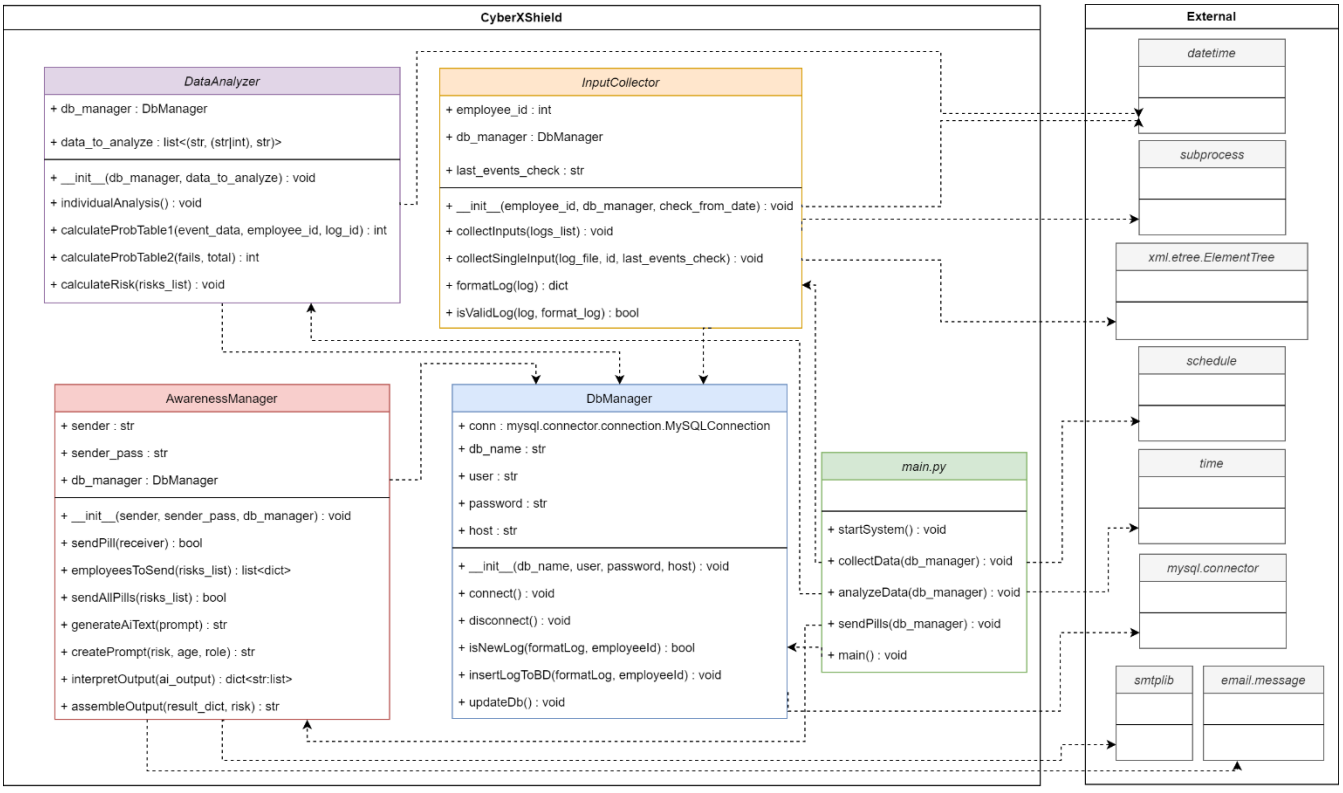


Fig 6. Diagrama de classes

A5. DASBOARD PER A LA VISUALITZACIÓ INTERACTIVA DE L'AVALUACIÓ DE RISCOS CREAT EN POWER BI

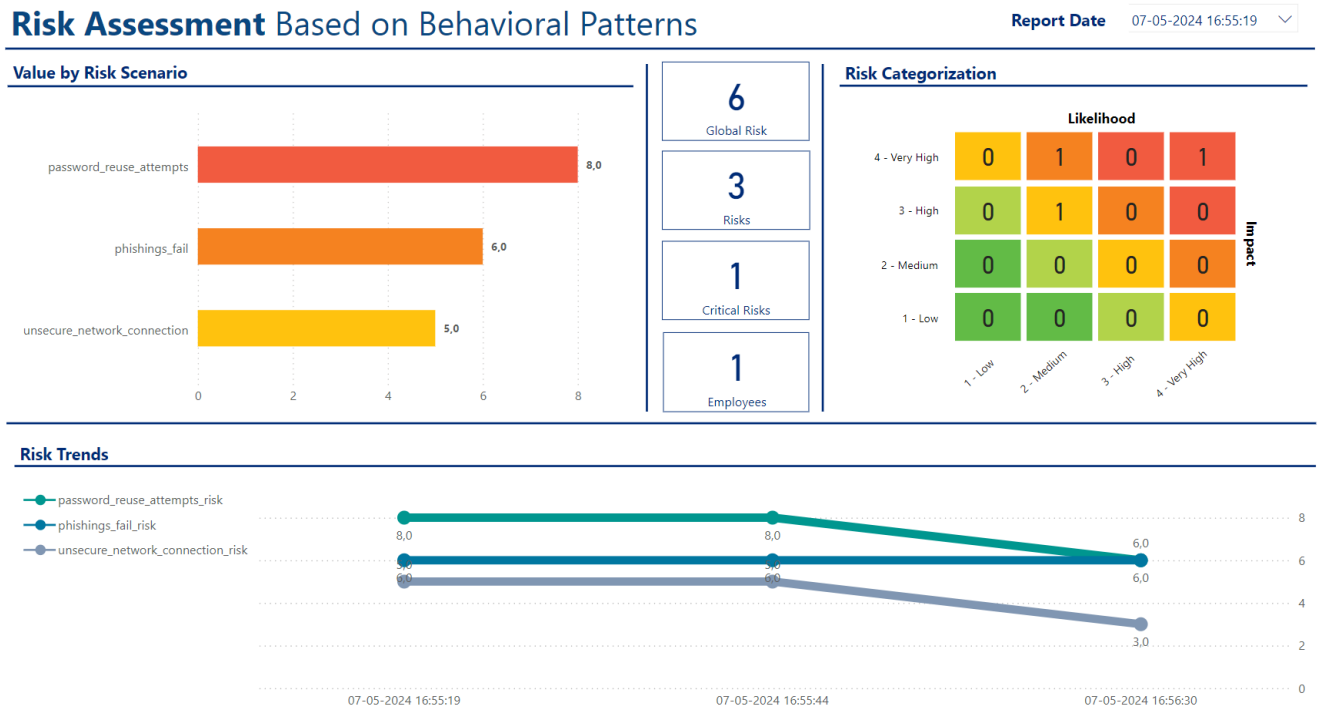
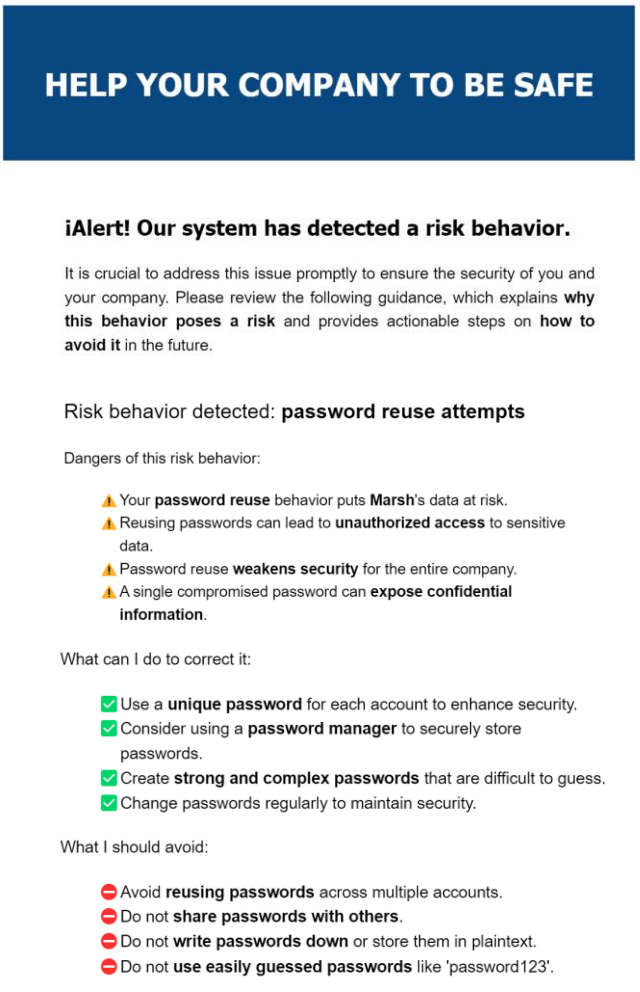


Fig 7. Dasboard per a la visualització interactiva de l'avaluació de riscos creat en Power BI

A6. RESULTATS DE LES PÍNDOLES GENERADES AMB INTEL·LIGÈNCIA ARTIFICIAL



HELP YOUR COMPANY TO BE SAFE

¡Alert! Our system has detected a risk behavior.

It is crucial to address this issue promptly to ensure the security of you and your company. Please review the following guidance, which explains **why this behavior poses a risk** and provides actionable steps on **how to avoid it** in the future.

Risk behavior detected: **password reuse attempts**

Dangers of this risk behavior:

- ⚠️ Your **password reuse** behavior puts **Marsh**'s data at risk.
- ⚠️ Reusing passwords can lead to **unauthorized access** to sensitive data.
- ⚠️ Password reuse **weakens security** for the entire company.
- ⚠️ A single compromised password can **expose confidential information**.

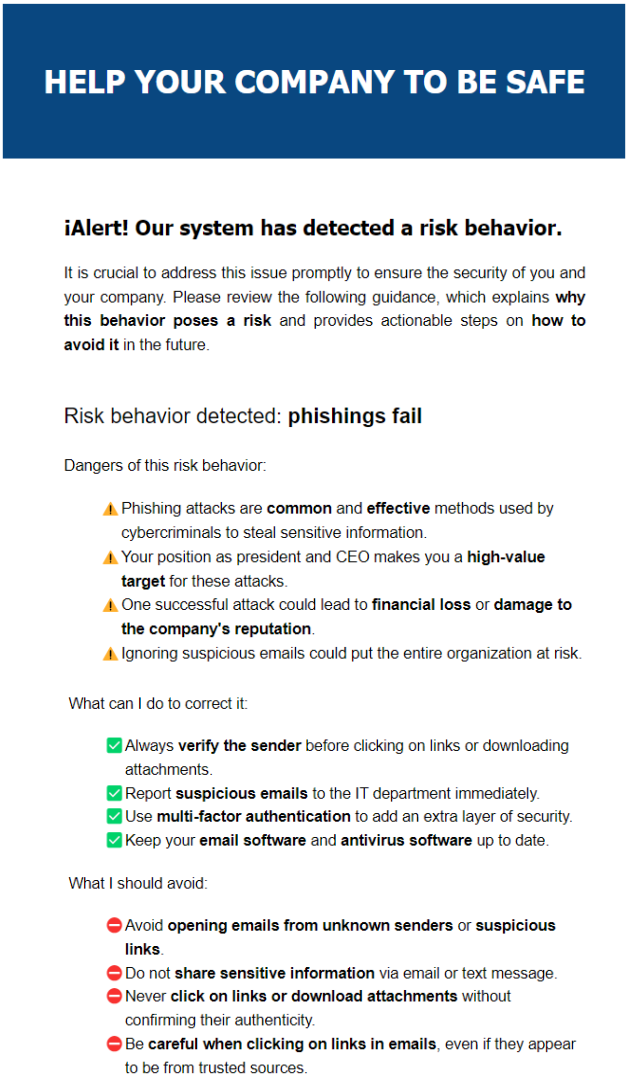
What can I do to correct it:

- ✅ Use a **unique password** for each account to enhance security.
- ✅ Consider using a **password manager** to securely store passwords.
- ✅ Create **strong and complex passwords** that are difficult to guess.
- ✅ Change passwords regularly to maintain security.

What I should avoid:

- ⛔ Avoid **reusing passwords** across multiple accounts.
- ⛔ Do not **share passwords with others**.
- ⛔ Do not **write passwords down** or store them in plaintext.
- ⛔ Do not **use easily guessed passwords** like 'password123'.

Fig 8. Píndola generada al detectar el comportament de risc "Reutilització de contraseñes"



HELP YOUR COMPANY TO BE SAFE

¡Alert! Our system has detected a risk behavior.

It is crucial to address this issue promptly to ensure the security of you and your company. Please review the following guidance, which explains **why this behavior poses a risk** and provides actionable steps on **how to avoid it** in the future.

Risk behavior detected: **phishings fail**

Dangers of this risk behavior:

- ⚠️ Phishing attacks are **common** and **effective** methods used by cybercriminals to steal sensitive information.
- ⚠️ Your position as president and CEO makes you a **high-value target** for these attacks.
- ⚠️ One successful attack could lead to **financial loss** or **damage to the company's reputation**.
- ⚠️ Ignoring suspicious emails could put the entire organization at risk.

What can I do to correct it:

- ✅ Always **verify the sender** before clicking on links or downloading attachments.
- ✅ Report **suspicious emails** to the IT department immediately.
- ✅ Use **multi-factor authentication** to add an extra layer of security.
- ✅ Keep your **email software** and **antivirus software** up to date.

What I should avoid:

- ⛔ Avoid **opening emails from unknown senders** or **suspicious links**.
- ⛔ Do not **share sensitive information** via email or text message.
- ⛔ Never **click on links or download attachments** without confirming their authenticity.
- ⛔ Be **careful when clicking on links in emails**, even if they appear to be from trusted sources.

Fig 9. Píndola generada al detectar el comportament de risc "Exercicis de phishing no superats"