Sistema de Información orientado para el manejo de datos sobre ataques de seguridad cibernética para el análisis en infraestructuras de red y servidores

Integrantes:

- **Alejandro Mollinedo**
- **Ronald Narvaez**
- **Samuel Sanjines**
- **Yara Quispe**
- **Rodrigo Rivera**

1. Introducción

En la actual era digital, los ataques cibernéticos han alcanzado un nivel de sofisticación y frecuencia sin precedentes, poniendo en riesgo la seguridad de infraestructuras críticas de red y servidores a nivel global. Las organizaciones enfrentan el desafío de identificar y mitigar estos riesgos de manera eficaz para proteger sus activos y datos. Este proyecto tiene como objetivo desarrollar un sistema de información especializado en el manejo y análisis de datos sobre ciberataques en infraestructuras de red y servidores, utilizando una combinación de tecnologías avanzadas que permitan detectar patrones, comportamientos y tendencias en los datos recopilados.

El sistema propuesto integrará herramientas de minería de datos, aprendizaje automático y visualización interactiva, presentando la información de forma gráfica a través de un dashboard accesible desde una página web. La plataforma permitirá analizar grandes volúmenes de datos globales sobre ciberataques, con el fin de proporcionar a los usuarios una comprensión más profunda y visual de las amenazas cibernéticas. Entre las tecnologías empleadas se incluyen Weka, Python, cloud computing, Power BI y Tableau. Estas herramientas facilitarán el procesamiento, análisis y representación gráfica de los datos, optimizando la toma de decisiones en relación con la seguridad de las redes y servidores.

El análisis de los datos se llevará a cabo mediante Weka, una plataforma que permite la aplicación de técnicas de clustering, regresión y clasificación, generando modelos predictivos para identificar patrones ocultos y tendencias en los ataques. Además, Python se utilizará para implementar algoritmos de aprendizaje automático y automatizar procesos de análisis de datos, mientras que el uso de tecnologías de cloud computing garantizará la escalabilidad y eficiencia en la gestión de grandes volúmenes de información.

El sistema web estará diseñado para ser accesible, interactivo y fácil de usar, proporcionando una plataforma intuitiva para la gestión y visualización de datos en tiempo real. Herramientas de visualización como Power BI o Tableau complementarán el análisis al integrar dashboards dinámicos que mostrarán los resultados de manera clara y concisa, facilitando la interpretación de la información por parte de los usuarios.

El análisis de datos se basará en un dataset de ciberataques a nivel mundial obtenido de Kaggle, una plataforma de competencia de ciencia de datos perteneciente a Google LLC. Este dataset, compuesto por 40.000 tuplas, incluye atributos como: Timestamp, Source IP Address, Destination IP Address, Source Port, Destination Port, Protocol, Packet Length, Packet Type, Traffic Type, Payload Data, Malware Indicators, Anomaly Scores, Alerts/Warnings, Attack Type, Attack Signature, Action Taken, Severity Level, User Information, Device Information, Network Segment, Geo-location Data, Proxy Information, Firewall Logs, IDS/IPS Alerts, y Log Source.

2. Lluvia de Ideas

a. Gráfico

b. Contexto

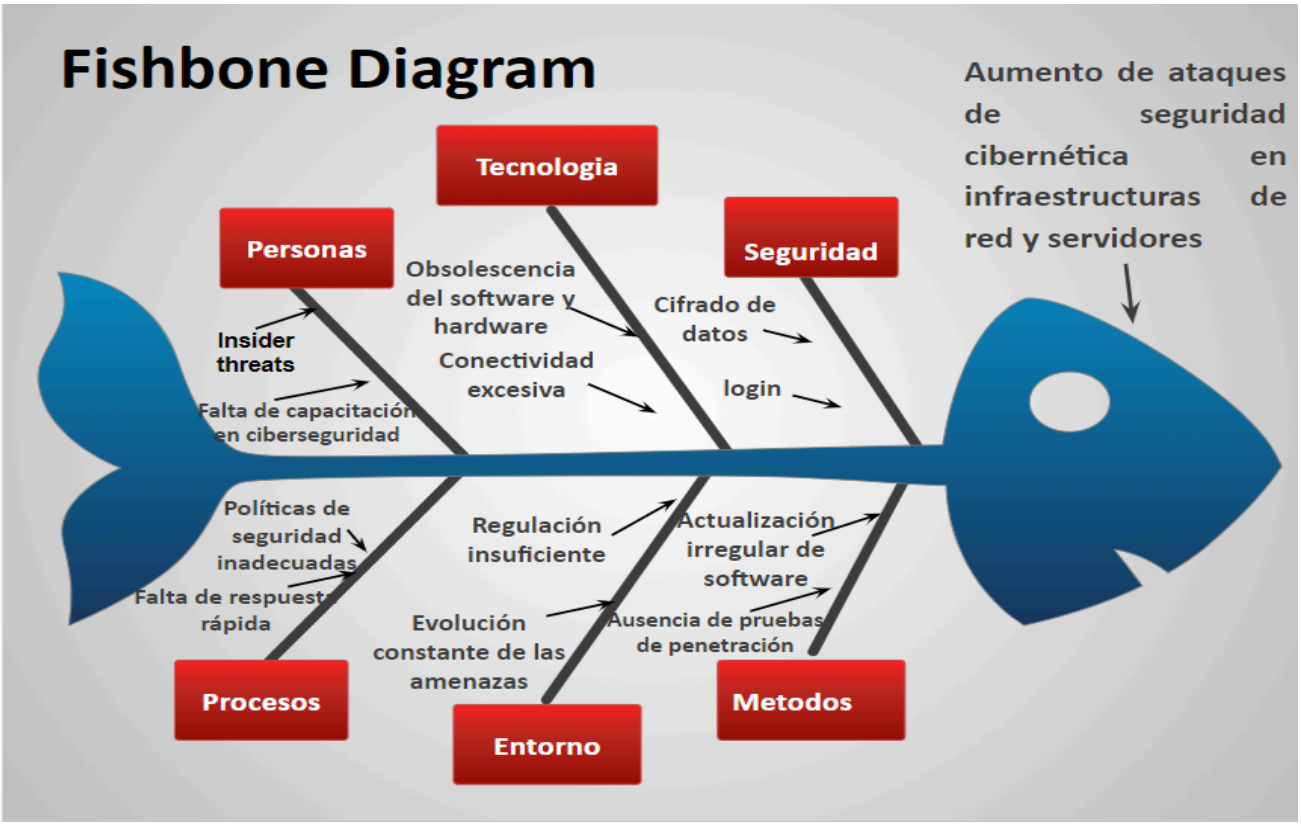
Hoy en día, los ataques cibernéticos se han vuelto un problema global debido al creciente uso de la tecnología y la interconectividad. Para analizar este fenómeno, trabajaremos con un dataset global que nos permitirá estudiar el comportamiento de estos ataques y cómo han evolucionado. Aplicaremos Cloud Computing para manejar el gran volumen de datos de forma adecuada y así garantizar la seguridad mediante sistemas de backup.

Usaremos herramientas como Power BI, Tableau, y Orange para crear dashboards interactivos que nos permitan visualizar los resultados y tendencias de manera más intuitiva. También aplicaremos técnicas de Data Mining y Machine Learning para identificar patrones y realizar predicciones, aprovechando el poder de tecnologías como Deep Learning.

Además, el uso de plataformas como Weka facilitará el análisis exploratorio de datos (EDA) para entender mejor los aspectos críticos de los ataques. Finalmente, la idea es integrar toda esta información en una página web que sea accesible y fácil de usar, combinando Business Intelligence con soluciones tecnológicas avanzadas.

3. Diagrama Ishikawa

a. Diagrama



b. Definición del problema

El problema identificado es el incremento de los ataques de seguridad cibernética en infraestructuras de red y servidores, impulsado por la creciente adopción de tecnologías digitales y la conectividad de dispositivos. Factores como la falta de actualización en los sistemas, políticas de seguridad inadecuadas, errores humanos y el uso de herramientas no confiables aumentan la vulnerabilidad frente a los ciberataques. Además, la evolución constante de las amenazas y la insuficiente regulación en materia de ciberseguridad complican la protección efectiva de los activos digitales, exponiendo a las organizaciones a riesgos significativos, como pérdidas de datos, interrupciones del servicio y daños a la reputación. Para abordar este desafío, se propone un enfoque integral que combine el análisis de datos de ciberataques mediante técnicas de minería de datos y aprendizaje automático para identificar patrones y mejorar las

defensas. El uso de tecnologías avanzadas como Cloud Computing para gestionar grandes volúmenes de datos, junto con herramientas de visualización y plataformas de análisis, permitirá explorar tendencias y reforzar la ciberseguridad de manera informada y efectiva.

4. Objetivo

a. General

Desarrollar un sistema de información orientado al análisis de datos sobre ataques de seguridad cibernética a nivel mundial, utilizando técnicas de minería de datos y aprendizaje automático para identificar patrones, tendencias y comportamientos en infraestructuras de red y servidores, con el fin de mejorar la ciberseguridad mediante visualizaciones gráficas interactivas y el manejo eficiente de grandes volúmenes de datos a través de tecnologías de business intelligence y Cloud Computing.

5. SMART

Específico (Specific)

- **¿Quién?** Nuestro equipo de desarrolladores de software como ser diseñadores y analistas de datos, junto con colaboradores especializados en análisis de datos y tecnologías Cloud.
- **¿Qué?** Desarrollar un sistema de información que permita analizar y visualizar datos globales de ataques cibernéticos, en el cual se utilizaran técnicas de minería de datos para identificar patrones, tendencias y comportamientos, mejorando la ciberseguridad en infraestructuras de red y servidores.
- **¿Dónde?** A nivel global, con un enfoque en infraestructuras de red y servidores.

- **¿Cuándo?** Inicio del proyecto en octubre, con un lanzamiento planificado para el final del semestre.
- **¿Cuáles?**
  - **Requerimientos:** Acceso a datasets globales de ciberataques, tecnologías de Cloud Computing, herramientas de visualización como Power BI o Tableau, y plataformas de análisis como Weka.
  - **Restricciones:** La complejidad de los algoritmos en la minería de datos podría generar altos tiempos de procesamiento, especialmente al manejar grandes volúmenes de datos. Además, la integración y compatibilidad de diversas herramientas de visualización y análisis que requieran un ajuste técnico detallado para garantizar un rendimiento óptimo del sistema.
- **¿Por qué?** Para mejorar la capacidad de defensa contra ciberataques, permitiendo a las organizaciones y gobiernos tomar decisiones basadas en datos sobre cómo proteger sus activos digitales.

Medible (Measurable)

- **¿Cuánto?** Un sistema funcional con análisis interactivo con la mayor cantidad de datos de los ciberataques registrados en el dataset con el que trabajaremos.
- **¿Cuántos?** Evaluar el rendimiento del sistema mediante la identificación de cierta cantidad de patrones relevantes de ataques cibernéticos y la creación de dashboards interactivos que visualicen las tendencias más importantes. Además, realizar pruebas de usabilidad con usuarios que utilicen el sistema para análisis de ciberseguridad.

- **¿Cómo saber si se ha logrado?** El éxito se medirá si el sistema genera visualizaciones interactivas con una gran precisión en la detección de patrones y anomalías en los ciberataques.

Alcanzable (Attainable)

- El proyecto es factible con las herramientas de análisis de datos y tecnologías de Cloud Computing disponibles actualmente. Los desarrolladores elaborarán un sistema eficiente y preciso.

Realista (Realistic)

- **¿Es posible?** Sí, el equipo cuenta con las habilidades y la experiencia necesarias en análisis de datos y conocimiento necesario sobre ciberseguridad y ataaques. Las tecnologías disponibles son suficientes para implementar el sistema propuesto.
- **¿Por qué?** El análisis de datos de ciberataques es un desafío crítico en la actualidad, y este proyecto ayudará a mejorar las defensas cibernéticas a través de la identificación de patrones y predicciones precisas.

Temporalizable (Timely)

- **¿Cuándo?** El sistema deberá estar completamente desarrollado y en funcionamiento antes de finalizar el semestre. Esto incluye el desarrollo y la implementación del sistema, así como la validación de los resultados mediante dashboards.

6. PART

Procesos:

- Desarrollo del sistema de información y análisis de ciberataques.
- Recopilación y procesamiento de datos globales de ciberataques.
- Aplicación de técnicas de minería de datos.
- Análisis y visualización de patrones y tendencias.

Actores:

- Analistas de datos.
- Expertos en ciberseguridad.
- Desarrolladores de software.
- Usuarios de plataformas de seguridad cibernética
- Administradores de infraestructuras de red y servidores.

Recursos:

- Datasets globales de ciberataques.
- Herramientas de visualización y análisis de datos (Power BI, Tableau, Orange, Weka).
- Tecnologías de Cloud Computing para el almacenamiento y análisis de grandes volúmenes de datos.
- Algoritmos de minería de datos y aprendizaje automático.

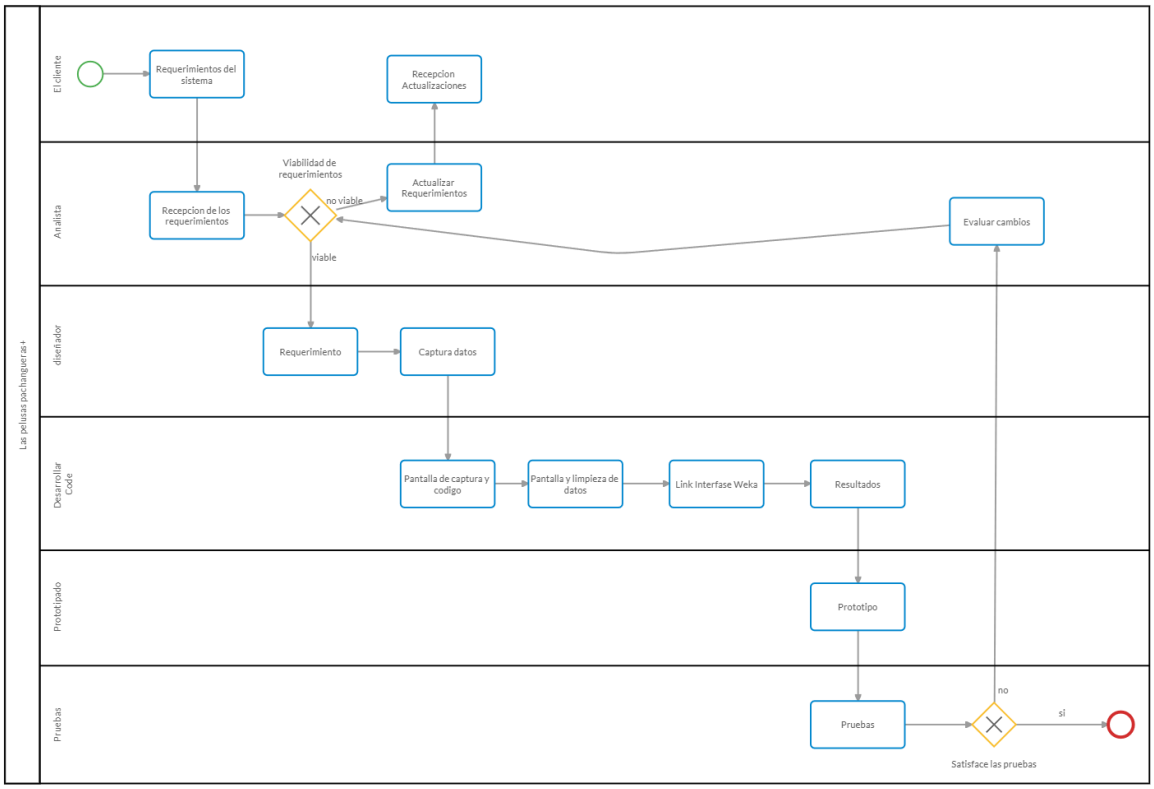
Tecnología:

- Cloud Computing.

- Herramientas de visualización interactiva (Power BI o Tableau).
- Plataformas de análisis de datos (Weka).

7. ANÁLISIS

Para realizar el análisis del desarrollo y el modelo de negocio del sistema web se utilizó BPMN (Business Process Model and Notation), que es un estándar gráfico que permite modelar y documentar procesos de negocio mediante diagramas de flujo, facilitando la comprensión y colaboración entre usuarios técnicos y no técnicos. Utiliza símbolos para representar eventos, actividades, decisiones y flujos de trabajo, lo que ayuda a optimizar procesos, identificar ineficiencias y sirve como base para la automatización. Su objetivo es proporcionar una representación clara y estandarizada de los procesos dentro de una organización. El diagrama BPMN se adjunta a continuación:



Title: Diagram
Author: Ronald Marcelo Narvaez Estevez
Export Date: 16/10/2024

El siguiente BPMN explica el proceso del modelo de negocio de nuestro sistema, comienza por el cliente enviando sus requerimientos, el analista del sistema es quien lo recibe, y tal como su labor indica analiza la viabilidad de estos, en caso de que estos requerimientos no cumplan con la viabilidad para ser un proyecto, se actualizará sus requerimientos a unos más viables, y el cliente recibirá esta actualización. En el caso de que sí sean viables, estos requerimientos ya definidos serán enviados al diseñador del sistema, el cual tendrá la tarea de recolectar datos para el proyecto, esta información (requerimientos y datos) serán enviados al desarrollador del sistema, el cual tiene la tarea de realizar la pantalla de captura y código, la pantalla y limpieza de datos, el link interface weka con todas las herramientas a usar que esta plataforma ofrece y finalmente mostrar los resultados.

En base a estos resultados, el encargado del prototipado procederá a realizar el prototipo del sistema y finalmente un beta-tester se encargará de realizar las pruebas necesarias para concluir con el funcionamiento del sistema. En caso de que estas pruebas no satisfacen el funcionamiento del sistema, el analista deberá evaluar los cambios en los requerimientos, y en el diseño del sistema. En caso de que estas pruebas logren satisfacer el funcionamiento del sistema, concluirá finalmente el desarrollo de este sistema.