DEPARTAMENTO DE CIENCIAS DE LA COMPUTACIÓN
CARRERA DE INGENIERÍA DE SOFTWARE

TRABAJO DE UNIDAD DE INTEGRACIÓN CURRICULAR, PREVIO A LA OBTENCIÓN DEL TÍTULO DE INGENIERO DE SOFTWARE

TEMA:

Implementación de un sistema de detección de intrusiones utilizando honeypots especializados en el procesamiento de pagos y generación de transacciones falsas, diseñados específicamente para su aplicación en entornos de transacciones en línea.

AUTORES:
JUMBO SALCEDO, KAREN LIZETH
GARCIA ROMERO, MARIO DARIO

DIRECTORA:
Ing. CORRAL DÍAZ, MARÍA ALEXANDRA, MSc

LATACUNGA FEBRERO, 2024

Orden del día

- Introducción
- Marco Teórico
- Implementación del Sistema
- Validación del Sistema
- Conclusiones

SECURE PAY GUARD

Orden del día

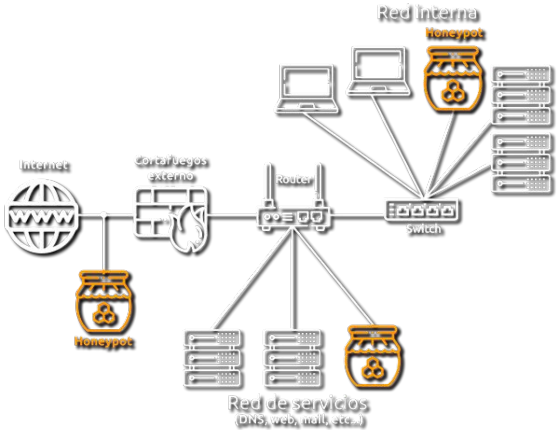
- Introducción
- Marco Teórico
- Implementación del Sistema
- Validación del Sistema
- Conclusiones

Problema

- Hoy en día, el aumento de las transacciones digitales ha llevado a un aumento significativo de las amenazas cibernéticas. Esto ha dado lugar a una creciente preocupación en el ámbito de la ciberseguridad.
- Estas transacciones en línea elevan el riesgo de ataques cibernéticos. Estas amenazas, que a menudo permanecen ocultas durante un período prolongado, permiten a los actores maliciosos acceder a información sensible sin ser detectadas

Planteamiento de la solución

- La importancia de este proyecto reside en su enfoque multifacético, centrándose en aspectos clave de la ciberseguridad y la protección de datos financieros en las transacciones en línea, busca ofrecer una solución proactiva y adaptativa para identificar las amenazas cibernéticas.
- El uso de técnicas de aprendizaje automático en este proyecto mejora significativamente la precisión de la detección, se adapta rápidamente a las nuevas amenazas, optimiza los recursos de seguridad, genera alertas tempranas y respuestas eficientes

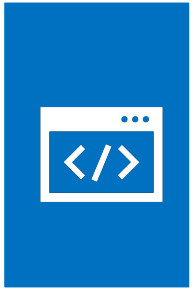


Objetivo General

Implementar un sistema de detección de intrusiones utilizando honeypots especializados en el procesamiento de pagos y generación de transacciones falsas, diseñados específicamente para su aplicación en entornos de transacciones en línea.

Objetivos Específicos

Comprender definiciones, terminología y el funcionamiento de los honeypots, así como los distintos tipos de honeypots diseñados para rastrear actividades en línea,



Implementar un honeypot especializado en el procesamiento de pagos y generación de transacciones falsas aplicado a transacciones online.

Validar resultados del sistema de detección de intrusos mediante honeypots aplicado a transacciones online.

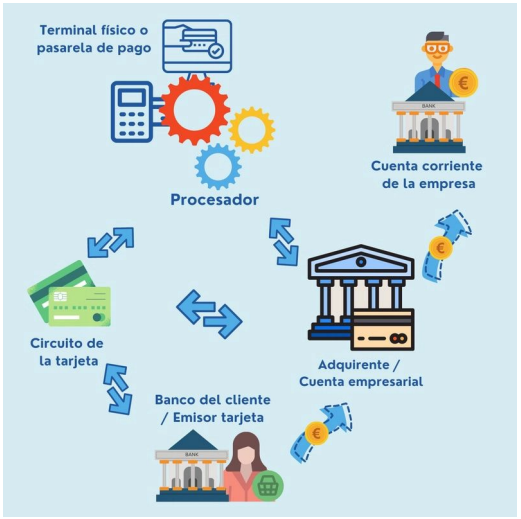
- Introducción
- Marco Teórico
- Implementación del Sistema
- Validación del Sistema
- Conclusiones

Sistemas Transaccionales en línea

- Están diseñados para transacciones comerciales rápidas y eficientes que utilizan tecnologías inteligentes para gestionar grandes volúmenes de datos y requisitos de alto rendimiento.
- Estos datos necesarios se introducen, confirman y autentican antes de la finalización (Internacional, 2021).
- Estos sistemas se caracterizan por transacciones cortas y frecuentes, grandes volúmenes de datos y altos requisitos de rendimiento, destacando su complejidad y la necesidad de protección contra intrusiones.

Procesamiento de pagos en línea

- El proceso de seguridad implica múltiples etapas desde la solicitud inicial hasta la confirmación y autorización por las instituciones financieras, centrándose en la seguridad a través de las normas PCI DSS y las tecnologías SSL/TLS.
- La observación sistemática y continua es crucial utilizando sistemas de detección de intrusiones (IDS) y sistemas de prevención de intrusión (IPS). (IPS).
- La detección de actividades anómalas se basa en el análisis de comportamientos inusuales, haciendo hincapié en la importancia en la detección y prevención de fraudes (Tariq et al., 2023), .

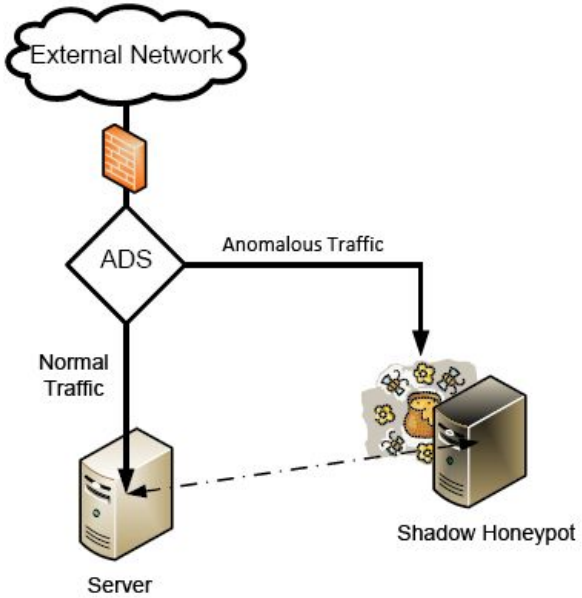


Suplantación de Identidad o Spoofing

- Spoofing es una técnica común en ciberataques, donde los delincuentes falsifican la identidad de remitentes o receptores para obtener información confidencial
- Hay varios métodos, como correo electrónico, páginas web y identificadores de llamadas (Proofpoint, 2024).
- Para prevenir el robo de identidad, se recomienda proteger y autenticar la dirección IP del cliente utilizando algoritmos criptográficos.

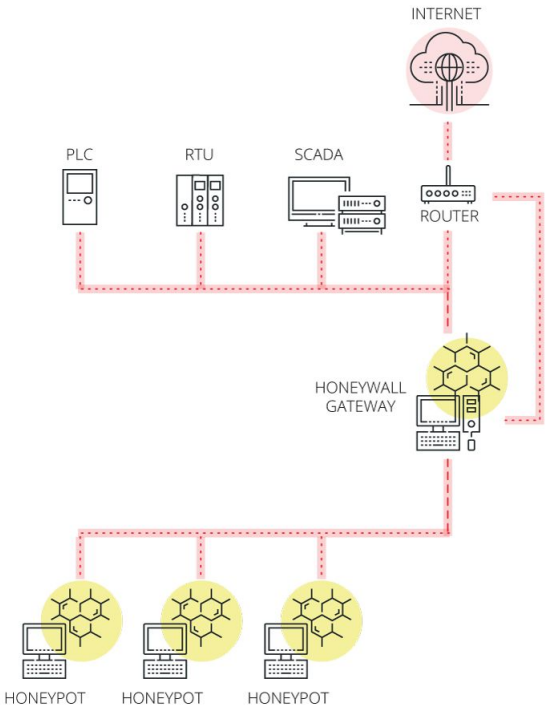
Honeypots

- Los honeypots son tecnologías activas de defensa que atraen a los atacantes estudiando sus actividades y métodos de aprendizaje, complementando los sistemas de seguridad pasiva (Tian et al., 2020).
- Vienen en varios tipos y niveles de interacción, cada uno diseñado para diferentes objetivos de seguridad y análisis de ataque.
- Los honeypots son especialmente útiles en los sistemas transaccionales, protegiendo las transacciones financieras y mejorando la ciberseguridad contra diversos métodos de ataque e intrusión.



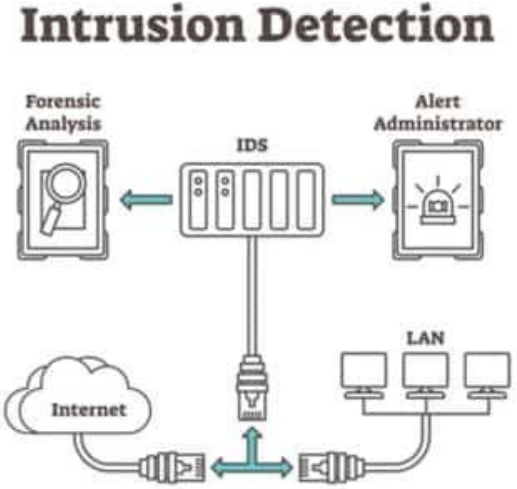
Honeypots para sistemas transaccionales

- Los honeypots para sistemas transaccionales son un tipo de honeypot que se utiliza con el fin de proteger los sistemas que procesan transacciones financieras.
- Estos sistemas suelen ser objetivos atractivos de los atacantes, ya que pueden obtener acceso a información confidencial o realizar transacciones fraudulentas (Jaramillo Ramos & Ospina Beltrán, 2019)..
- Al simular ser sistemas legítimos, los honeypots para sistemas transaccionales pueden detectar y analizar posibles amenazas antes de que causen daño real.



Sistema de detección de intrusiones (IDS)

- Un IDS monitoriza y analiza la información que fluye a través de una red para identificar posibles ataques, cerrando las puertas al potencial intruso y reconfigurando elementos de la red como firewalls y routers.
- Los IDS son una herramienta fundamental en la detección y prevención de intrusiones, permitiendo una respuesta rápida y efectiva ante posibles amenazas cibernéticas.
- Su capacidad para identificar comportamientos anómalos en tiempo real es crucial para mantener la seguridad de la red y proteger la información sensible (Gómez López, 2009).



Algoritmos y/o modelos de Machine Learning

Random Forest

Es un método de aprendizaje supervisado que construye árboles de decisión durante el entrenamiento y hace predicciones sobre la base de las predicaciones de cada árbol. Utiliza múltiples modelos para mejorar la robustez y el rendimiento (Kazemian & Shrestha, 2023).

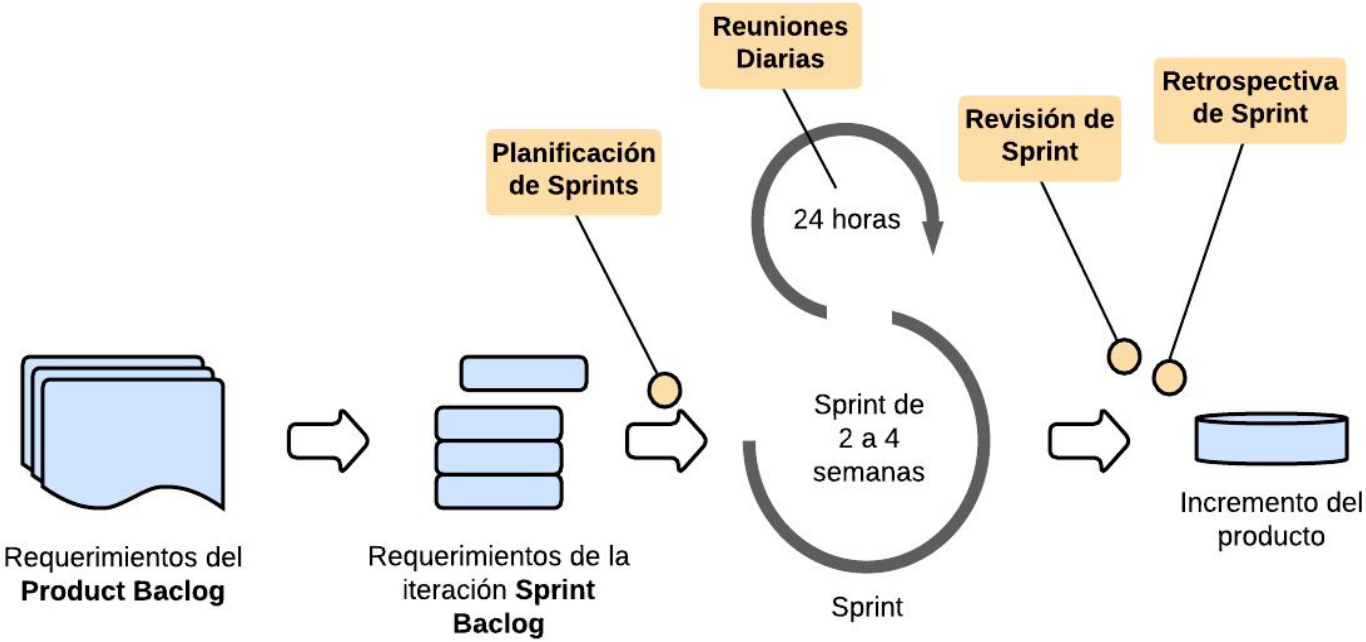
Regresión Logistica

La regresión logística es una técnica que analiza la relación entre una variable dependiente dicotómica y uno o más predictores independientes. Para fines explicativos y predictivos, puede manejar variables cuantitativas y categóricas (Kazemian & Shrestha, 2023).

- Introducción
- Marco Teórico
- Implementación del Sistema
- Validación del Sistema
- Conclusiones

Metodología de desarrollo

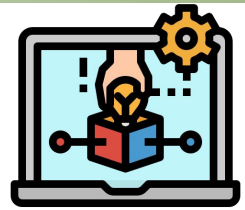
- Esquema de la metodología Scrum

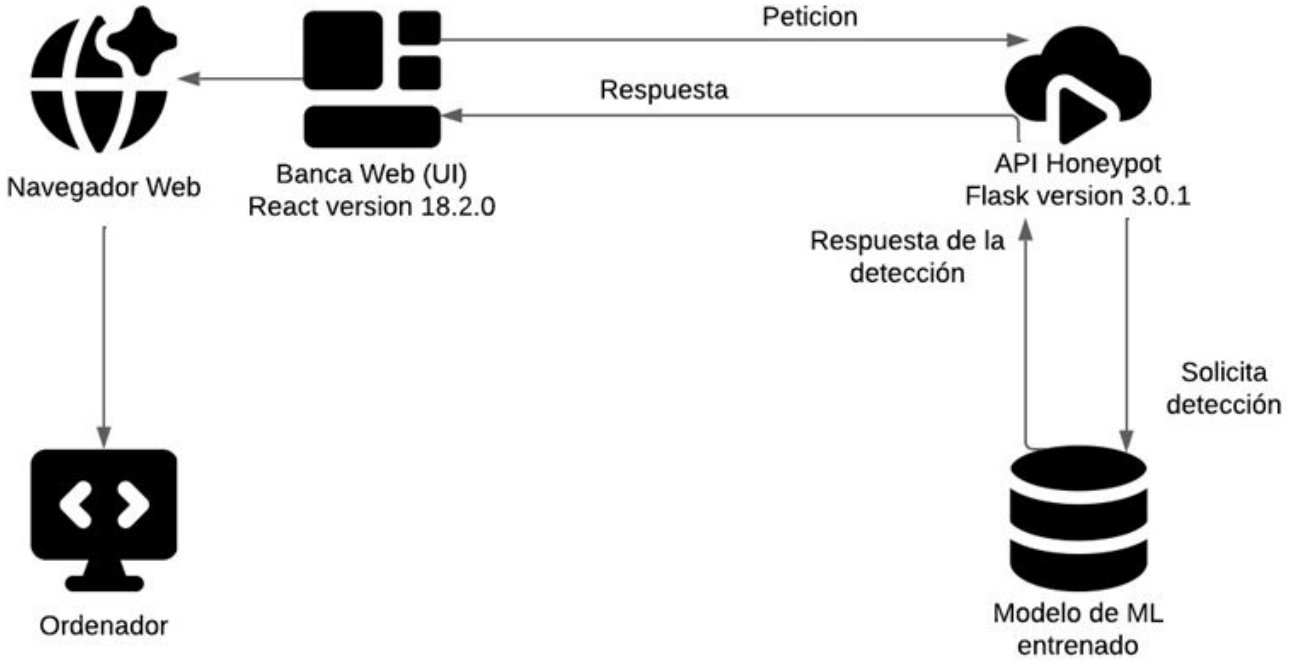


Recuperado de (Zayat & Senvar, 2020)

Diseño del sistema

- Arquitectura Física del sistema





Diseño del sistema

Sprint 01: Creación del dataset, modelo de Machine Learning e Implementación del Honeypot

Características del usuario por transacción

Característica	Detalle
Propietario	Identificación del titular de cuenta implicada en la transacción.
Servicio	Especifica el tipo de servicio: Luz, agua, teléfono, internet
Código de pago	Identificador único, asignado a cada transacción para cada pago de servicio realizado
Cédula	Número de identificación del propietario de la cuenta
Fecha / Hora de pago	Marca la actividad temporal de la transacción
Dirección IP	Dirección del dispositivo utilizado en Internet
Monto	La cantidad de dinero involucrada para cada transacción de pago realizado por el usuario
Estado	Detecta el estado de la transacción que puede ser, normal o anómala.

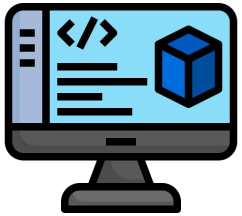
Dataset

```python
# Crear un DataFrame con los datos de transacciones
df = pd.DataFrame(transacciones,
                  columns=['propietario', 'servicio', 'codigo de pago', 'cedula',
                           'fecha_hora_de_pago', 'direccion_ip', 'estado', 'monto'])

# Guardar el DataFrame en archivos CSV separados para entrenamiento y prueba
df_entrenamiento = df.sample(frac=0.8, random_state=42)
df_prueba = df.drop(df_entrenamiento.index)

df_entrenamiento.to_csv('transacciones_entrenamiento.csv', index=False)
df_prueba.to_csv('transacciones_prueba.csv', index=False)
```

```
_id: ObjectId('65d0c84e7bf7ed0d4166916f')
servicio: "Luz"
monto: 19
codigo_pago: "maria_luz_864"
nombre: "maria"
cedula: "1234567891"
fecha_pago: "2024-02-17T14:53:01.483Z"
ip_pago: "190.99.77.147"
```

Diseño del sistema

Sprint 01: Creación del dataset, modelo de Machine Learning e Implementación del Honeypot



Entrenamiento del Honeypot

```
# Seleccionar caracteristicas y etiquetas
X = df_entrenamiento[['hora', 'monto']]
y = df_entrenamiento['estado']

# Dividir los datos en conjunto de entrenamiento y prueba
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

# Inicializar y entrenar el modelo de Random Forest
random_forest_model = RandomForestClassifier(n_estimators=100, random_state=42)
random_forest_model.fit(X_train, y_train)

# Inicializar y entrenar el modelo de Regresion Logistica
logistic_regression_model = LogisticRegression(random_state=42)
logistic_regression_model.fit(X_train, y_train)

# Predecir las etiquetas en el conjunto de prueba
y_pred_random_forest = random_forest_model.predict(X_test)
y_pred_logistic_regression = logistic_regression_model.predict(X_test)
```

Cálculo de Métricas

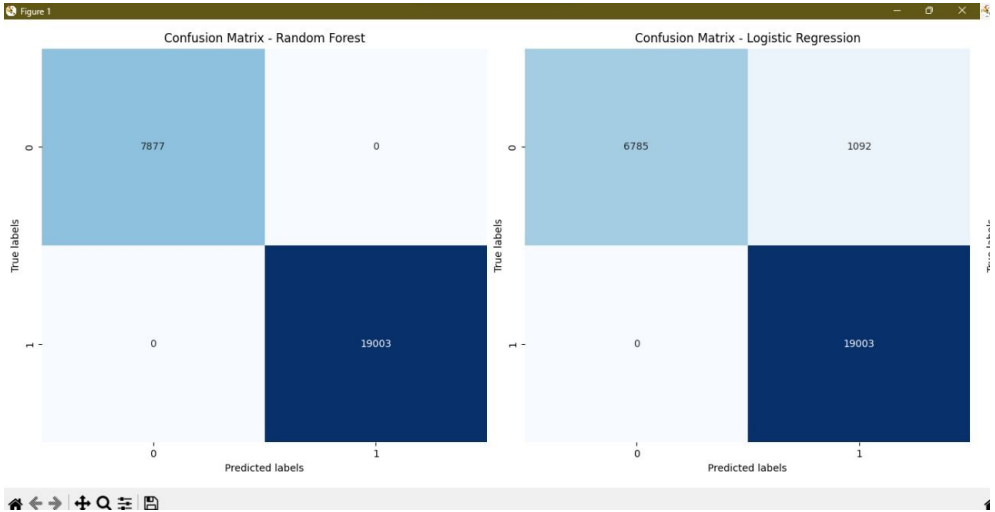
```
# Calcular metricas para Random Forest
accuracy_rf = accuracy_score(y_test, y_pred_random_forest)
precision_rf = precision_score(y_test, y_pred_random_forest, pos_label='anomalo')
recall_rf = recall_score(y_test, y_pred_random_forest, pos_label='anomalo')
confusion_matrix_rf = confusion_matrix(y_test, y_pred_random_forest)

# Calcular metricas para Regresion Logistica
accuracy_lr = accuracy_score(y_test, y_pred_logistic_regression)
precision_lr = precision_score(y_test, y_pred_logistic_regression, pos_label='anomalo')
recall_lr = recall_score(y_test, y_pred_logistic_regression, pos_label='anomalo')
confusion_matrix_lr = confusion_matrix(y_test, y_pred_logistic_regression)
```

Diseño del sistema

Sprint 01: Creación del dataset, modelo de Machine Learning e Implementación del Honeypot

Métricas y Matriz de confusión

```
Random Forest:
Accuracy: 0.98679
Precision: 1.0000
Recall: 0.99679

Logistic Regression:
Accuracy: 0.9389
Precision: 1.000
Recall: 0.8476

Process finished with exit code 0
```



Elaboración propia

Desarrollo del Sistema

Sprint 02. Desarrollo de Sistema bancario simulado

SECUREBANK

Seleccione una opción:

Iniciar Sesión | Registrarse | Mi Banco Web

Inicio de Sesión Banco

Correo electrónico:

Contraseña:

Iniciar Sesión

¿No tienes una cuenta? Regístrate

Página de Servicios

Seleccione un servicio

Luz | Agua | Teléfono | Internet

Has seleccionado Luz. Ingresa los datos del usuario:

Nombre:

Cédula:

El monto a pagar es de $19

Ingresa el código de pago:

Pagar

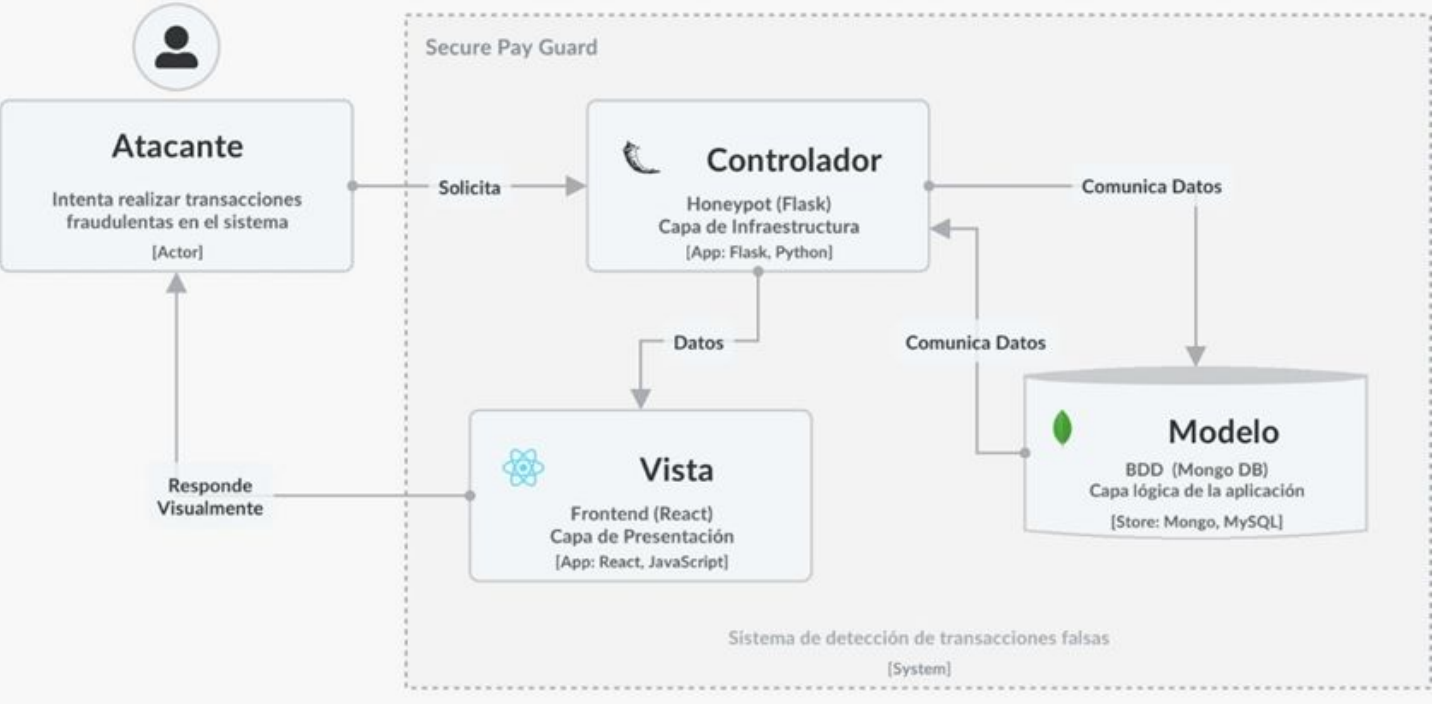
localhost:3000 dice

Pago procesado correctamente

Aceptar

Desarrollo del Sistema

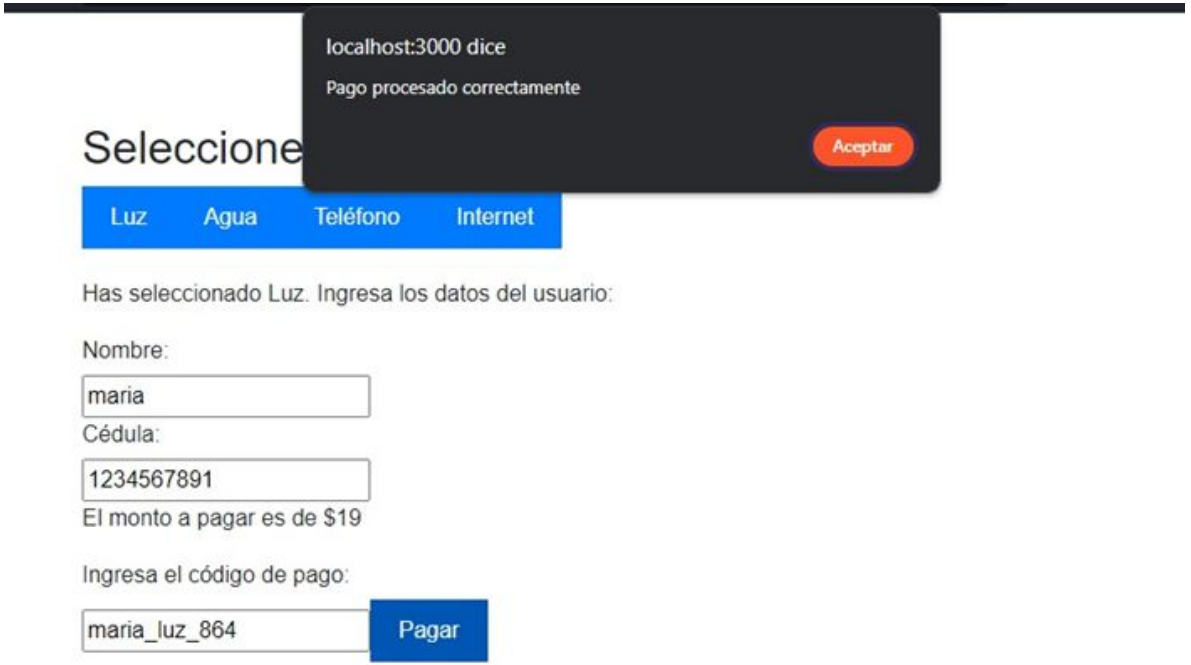
Arquitectura del modelo



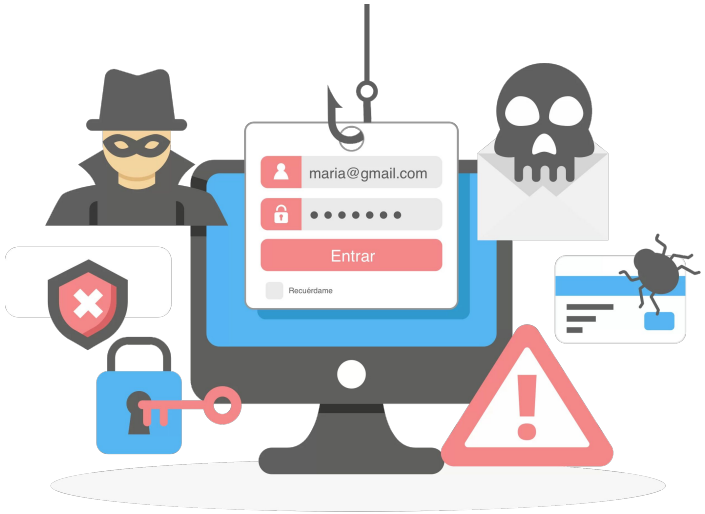
Elaboración propia

Desarrollo del Sistema

- Ejecución del sistema



- Introducción
- Marco Teórico
- Implementación del Sistema
- Validación del Sistema
- Conclusiones



Validación del Sistema

- Se describe el uso de la validación cruzada para evaluar el rendimiento del modelo con diferentes conjuntos de datos y se destacan las métricas de precisión, exactitud y sensibilidad como criterios de evaluación.
- Se presentan escenarios de prueba específicos, enfocados a ajustes en la selección de características, además de la implementación de modelos como la regresión logística y Random Forest.

Segmentación de grupos simulados

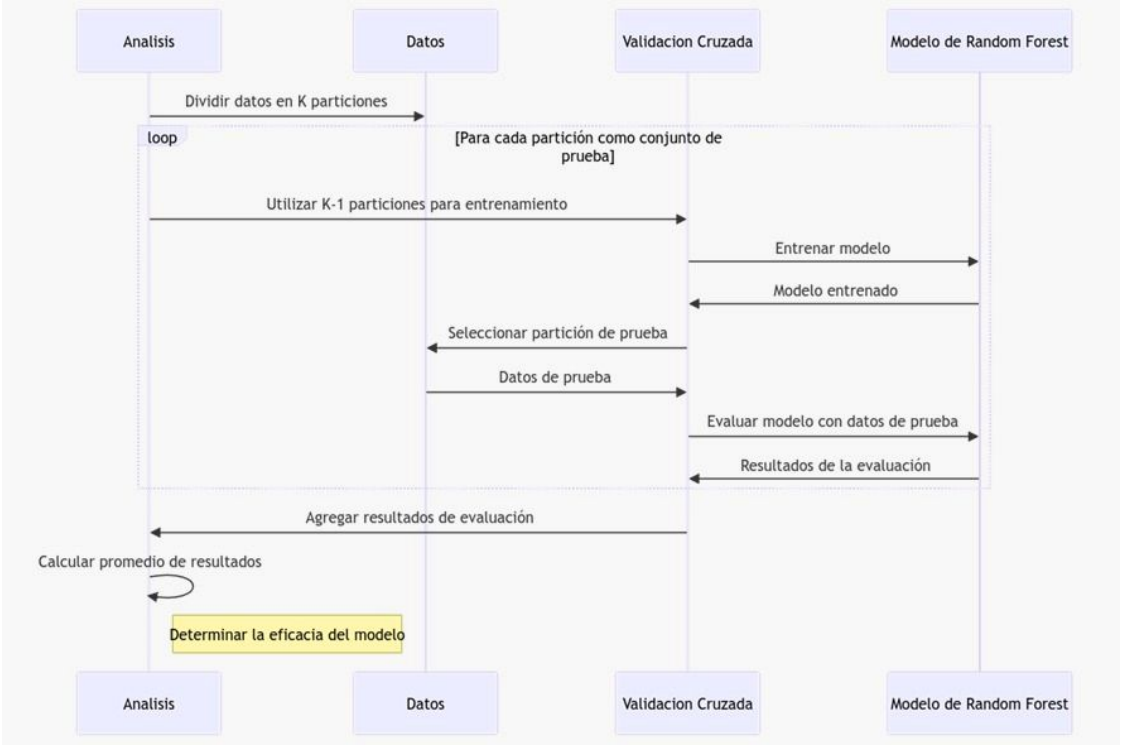
Pagos Mensuales Aproximados	Bajos Recursos (USD)	Medios Recursos (USD)	Altos Recursos (USD)
Luz	$20 - $40	$30 - $50	$50 - $80
Agua	$10 - $20	$15 - $30	$30 - $50
Teléfono	$10 - $30	$20 - $40	$30 - $60
Internet	$20 - $50	$30 - $70	$50 - $100

Nota: Pagos Mensuales Aproximados analizados de informes oficiales del INEC para el año 2023

Elaboración propia

Validación del Sistema

- Proceso de validación cruzada



Validación del Sistema

- Obtención de datos para validar el sistema

Matriz de confusión

	Positivos	Negativos
Positivos	Spoofing clasificados correctamente (VP)	Legítimos mal clasificados (FP)
Negativos	Spoofing mal clasificado (FN)	Legítimos clasificados correctamente (VN)

Nota. Formato para evaluación del sistema.

Métricas de evaluación

Métricas	Formula
Accuracy	$acurracy = \frac{VP + VN}{VP + VN + FP + FN}$
Recall	$recall = \frac{VP}{VP + FN}$
Precisión	$precision = \frac{VP}{VP + FP}$

Validación del Sistema

- Obtención de las métricas de evaluación en los 2 escenarios

Resultados de pruebas del modelo y algoritmo implementado primer escenario

Ord	Características	Algoritmos/Modelos	Accuracy	Recall	Precision
1	25 000	Random Forest	0.98679	1.0000	0.99679
2		Regresión Logística	0.9389	1.0000	0.8476

Nota. Resultados primer escenario con las métricas de eva

Machine Learning.

Resultados de pruebas del modelo y algoritmo implementado segundo escenario

Ord	Características	Algoritmos/Modelos	Accuracy	Recall	Precision
1	40 000	Random Forest	1.0000	1.0000	1.0000
2		Regresión Logística	0.9594	1.0000	0.8614

Nota. Resultados segundo escenario con las métricas de evaluación a los modelos de

Machine Learning.

Análisis de resultados

- Durante la fase de validación de Secure Pay Guard, se llevó a cabo un análisis comparativo de dos modelos de Machine Learning. Los resultados mostraron que el modelo de Regresión Logística mostró una disminución en la precisión, lo que indica una tendencia a identificar incorrectamente las transacciones legítimas como fraudulentas. La extensión del análisis a un conjunto de datos más grande mostró que Random Forest mantuvo su rendimiento excepcional, mejorando todas las métricas evaluadas a una puntuación perfecta. Sin embargo, el modelo de Regresión Logística se enfrentaba a limitaciones, con una precisión del 86,14%, lo que indica dificultades para manejar la creciente complejidad. Los resultados muestran la importancia de un enfoque iterativo y contextual para desarrollar sistemas que puedan encontrar anomalías.

- Introducción
- Marco Teórico
- Implementación del Sistema
- Validación del Sistema
- Conclusiones

Conclusiones

- Se ha logrado una profunda comprensión de las definiciones, la terminología y el funcionamiento de los honeypots, cumpliendo el primer objetivo específico. Este estudio ha examinado varios tipos de honeypots, ofreciendo una comprensión detallada del proceso de desarrollo. La investigación sobre técnicas computacionales avanzadas ha permitido la implementación de un honeypots especializado, destacando la importancia de una base teórica sólida encaminando a el éxito práctico del sistema.

Conclusiones

- La adopción de una arquitectura en capas y el patrón MVC se convierten en elementos indispensables relacionados a el diseño y la gestión de sistemas seguros y escalables, como lo requiere el sistema de detección de intrusos en entornos transaccionales.
- La puesta en práctica de un honeypot especializado en el procesamiento de pagos y la generación de transacciones falsas, se ha cumplido con éxito. Este avance ha permitido la implementación directa del sistema en las transacciones en línea, destacando su capacidad con el fin simular vulnerabilidades, facilitando así su detección y análisis.

Conclusiones

- La eficacia de este enfoque especializado en el entorno de las transacciones en línea subraya la importancia de adaptar las herramientas de seguridad a las necesidades específicas del contexto en el que se aplicarán.
- Las pruebas realizadas en entornos simulados han confirmado la capacidad del sistema para detectar y registrar con precisión intentos de intrusión. Esta validación no sólo demuestra la funcionalidad del sistema en diversas condiciones, sino que también asegura su aplicabilidad y fiabilidad en entornos reales de transacciones en línea.

Recomendaciones

- Es importante buscar palabras claves que se relacionen con el tema de investigación, con el fin encontrar información más relevante y útil de cara a empezar tu investigación.
- La ampliación del conjunto de datos para entrenar modelos de Machine Learning es crucial, enfocando una mayor variedad de características que reflejan los patrones de comportamiento del usuario en escenarios de transacciones reales, enfocando en factores psicológicos, contextuales y tecnológicos que pueden influir en las transacciones en línea.

Recomendaciones

- Es crucial implementar un proceso de evaluación y ajuste continuos de los modelos de aprendizaje automático utilizados. Este proceso debería implicar la validación periódica de la eficacia del modelo con nuevos conjuntos de datos, la reevaluación de las características utilizadas y la exploración de nuevos modelos o técnicas de aprendizaje automático que puedan proporcionar mejores resultados en la detección de intrusiones.
- La adaptabilidad y la mejora continua son esenciales hacia mantener la eficacia del sistema frente a la evolución de las tácticas de intrusión y los patrones de comportamiento del usuario.

Bibliografía

- Internacional. (2021, agosto 17). ¿Cómo realizar una transferencia bancaria? Banco Internacional. https://www.bancointernacional.com.ec/como-realizar-una-transferencia-bancaria/
- Tariq, E., Akour, I., Al-shanableh, N., Alquqa, E., Alzboun, N., Ibra-Heem, S., Al-Hawary, S., & Alshurideh, M. (2023). How cybersecurity influences fraud prevention: An empirical study on Jordanian commercial banks. International Journal of Data and Network Science, 8. https://doi.org/10.5267/j.ijdns.2023.10.016

Bibliografía

- Proofpoint. (2024). ¿Qué es el robo de identidad? - Definición, ejemplos y tipos | Proofpoint ES. https://www.proofpoint.com/es/threat-reference/identity-theft
- Kazemian, H., & Shrestha, S. (2023). Comparisons of machine learning techniques for detecting fraudulent criminal identities. Expert Systems with Applications, 229, 120591. https://doi.org/10.1016/j.eswa.2023.120591
- Gómez López, J. (2009). Optimización de sistemas de detección de intrusos en red utilizando técnicas computacionales avanzadas [Http://purl.org/dc/dcmitype/Text, Universidad de Almería]. https://dialnet.unirioja.es/servlet/tesis?codigo=22175

Bibliografía

- Tian, W., Ji, X., Liu, W., Liu, G., Zhai, J., Dai, Y., & Huang, S. (2020). Prospect Theoretic Study of Honeypot Defense Against Advanced Persistent Threats in Power Grid. IEEE Access, 8, 64075-64085. https://doi.org/10.1109/ACCESS.2020.2984795
- Jaramillo Ramos, C. C., & Ospina Beltrán, M. I. (2019). Arquitectura de integración basada en tecnología Blockchain para sistemas transaccionales con bases de datos distribuidas. Caso de estudio: Facturación agencia de viaje. http://repository.udistrital.edu.co/handle/11349/22898

Software
DOMINA EL MUNDO

Gracias por su atención

ESPE
UNIVERSIDAD DE LAS FUERZAS ARMADAS
INNOVACIÓN PARA LA EXCELENCIA