Laurea Triennale in informatica - Università di Salerno
Corso di *Ingegneria del Software* - Prof.ssa F. Ferrucci

GUARDIAN FLOW

FEELING SAFE

Manuale di Sistema

Guardian Flow

Riferimento	
Versione	0.2
Data	
Destinatario	Prof.ssa F. Ferrucci
Presentato da	Vincenzo Maiellaro
Approvato da	Raffaele Mezza, Martina Mingione

Sommario

Revision History 3
1. Introduzione 4
2. Tecnologie Utilizzate 5
2.1 Front-End (Nuxt.js) 5
2.2 Back-End (Python) 5
2.3 Database (PostgreSQL) 5
3. Intelligenza Artificiale non supervisionata 6
4. Ricezione del Traffico 7
5. Monitoraggio e Manutenzione 8
5.1 Monitoraggio Anomalie 8
5.2 Aggiornamenti Software 8
6. Risoluzione Problemi 8
7. Sicurezza 8
8. Glossario 9

Revision History

Data	Versione	Descrizione	Autori
03/01/2024	0.1	Stesura prima parte	Vincenzo Maiellaro
04/01/2024	0.2	Stesura seconda parte	Vincenzo Maiellaro
04/01/2024	0.2	Revisione	Vincenzo Maiellaro
04/01/2024	0.2	Approvazione	Martina Mignone Raffaele Mezza

1. Introduzione

Guardian Flow è il risultato della necessità di creare una solida struttura digitale per analizzare il traffico di rete di un'azienda alla ricerca di eventuali anomalie. Tra gli obiettivi vi è quello di migliorare il processo esistente, il quale ha evidenziato problematiche legate a una rilevazione inefficace delle anomalie nel traffico di rete, spingendoci ad adottare un approccio più avanzato e scalabile. Guardian Flow costruisce una baseline personalizzata per ogni cliente e mette a disposizione una dashboard che consente ai clienti di monitorare in tempo reale il traffico di rete analizzato e le eventuali anomalie individuate. Inoltre, il sistema offre la possibilità agli utenti di modificare agevolmente il piano di abbonamento, garantendo loro la libertà di adattare il sistema e le risorse disponibili alle specifiche esigenze aziendali.

L'obiettivo di questo documento è di rendere quanto più chiaro possibile, e in poche parole, la struttura del sistema e il suo funzionamento andando anche a spiegare in breve le tecnologie utilizzate indicando i motivi per cui si è scelto di utilizzarle a scapito di altre soluzioni.

2. Tecnologie Utilizzate

2.1 Front-End (Nuxt.js)

Nuxt.js è un framework basato su Vue.js che semplifica lo sviluppo di applicazioni web a pagina singola e di rendering lato server.

Una particolarità di Nuxt.js è l'ottimizzazione delle prestazioni, cosa di estrema importanza nel caso di un sistema di questo tipo in quanto esso deve essere molto reattivo e bisogna usare al meglio le risorse a disposizione, quindi, ridurre il carico della parte front-end lascerà alla parte back-end delle risorse extra, il che è vitale in casi dove c'è un sovraccarico di traffico inoltrato e le risorse richieste sono molte.

2.2 Back-End (Python)

La scelta del linguaggio di programmazione del lato back-end dell'applicativo è ricaduta su Python.

Molti applicativi basati su Intelligenza Artificiale, siano essi supervisionati o meno, sono scritti proprio in Python, quindi, è un linguaggio già "formato" a questo scopo, anche grazie a molte librerie atte proprio a questo, alcuni esempi sono TensorFlow, PyTorch e Scikit-learn.

Altro suo punto di forza è la chiarezza e semplicità, cosa che permette a team, anche voluminosi, di coordinarsi meglio nello sviluppo e nella modifica del codice.

2.3 Database (PostgreSQL)

Data l'enorme importanza che ha la sicurezza dei dati di accesso dei clienti e il salvataggio rapido e preciso delle anomalie, il database che si utilizza in questo sistema è PostgreSQL.

PostgreSQL è un database relazionale che esiste dal 1996 ed è tutt'ora molto utilizzato, tra le sue caratteristiche principali troviamo:

- Affidabilità dovuta alle transazioni ACID, le quali faranno si che i dati non possano sparire o diventare incoerenti non rispettando più i vincoli relazionali;
- Scalabilità e Flessibilità, il database offre molte opzioni per la scalabilità, quindi è preposto all'ottimizzazione su richiesta;
- Open Source, essendo Open Source, non ha costi di licenza.

3. Intelligenza Artificiale non supervisionata

L'IA non supervisionata è progettata per apprendere dati e schemi senza la necessità che venga guidata da un umano. A differenza della sua controparte supervisionata, in cui il modello riceve dati di input etichettati con output desiderati, l'apprendimento non supervisionato implica l'elaborazione di dati senza istruzioni esplicite.

L'uso dell'IA non supervisionata fornisce le seguenti caratteristiche importanti:

- Rilevazione delle anomalie sconosciute, definite comunemente come "0 Day Threat", cioè anomalie la cui esistenza è diventata nota da pochissimo. La caratteristica dell'algoritmo utilizzato da Guardian Flow, chiamato Isolation Forest, è la velocità con cui esso è in grado di identificare le anomalie, anche le 0 Day, senza dipendere esclusivamente dalle somiglianze con casi noti;
- Grandi capacità di adattamento, l'IA Non Supervisionata è capace di adattarsi, da sola e gradualmente, ai cambi di "normalità", è in grado di riconoscere cambiamenti nei pattern delle anomalie senza alcun input o riaddestramento. Il riaddestramento sarà necessario, invece, se la struttura della rete del cliente cambia in modo significativo;
- Individuazione di pattern complessi, nelle reti di grandi dimensioni e nei flussi di dati complessi, quindi anche un traffico congestionato, l'IA Non Supervisionata può individuare pattern che potrebbero essere difficilmente identificabili da regole predefinite, data la sua flessibilità.

4. Ricezione del Traffico

Successivamente all'acquisto e prima dell'attivazione del sistema e l'invio del traffico, il cliente deve fornire un buon volume di dati di esempio su cui costruire la sua baseline, essa sarà la base su cui l'IA andrà ad effettuare le sue valutazioni nelle prime istanze di analisi del traffico ricevuto. Proprio per questo è di vitale importanza che questi dati di esempio siano corretti e variegati per permettere all'IA di rilevare più tipi di anomalie e, soprattutto, è importante che in NESSUNA istanza di questi dati di esempio ci siano anomalie, il motivo per cui è essenziale che non ci siano è perché l'IA, andando ad addestrarsi sulle anomalie, non le rileverà come minacce quando l'applicativo è in funzione.

Per l'abilitazione del sistema, una volta costruita la baseline, al cliente verrà fornito un numero di porta sulla quale mandare il traffico che vuole far esaminare.

Il traffico che il cliente genera verrà inoltrato tramite protocollo SFTP e, per restare conformi al GDPR, la fonte del traffico non sarà nota.

Dopo ciò, il traffico verrà elaborato in Python (tramite i dovuti Parser json) e successivamente verrà analizzato dall'intelligenza artificiale, la quale sarà in grado di classificarlo appropriatamente.

5. Monitoraggio e Manutenzione

5.1 Monitoraggio Anomalie

Tramite la Dashboard, un utente può accedere facilmente a tutte le anomalie rilevate dal sistema, andando ad esaminarne una in particolare, sarà possibile vederne i dettagli, il livello di criticità e, se l'utente è Amministratore, segnalarla come un Falso Positivo, andando anche a dare un riscontro diretto all'IA che ne terrà conto nelle prossime analisi.

5.2 Aggiornamenti Software

Poiché si tratta di un sistema basato sull'IA Non Supervisionata, gli aggiornamenti del sistema si basano principalmente sull'auto-miglioramento dell'IA stessa, man mano che essa fa il suo lavoro. Il cliente, quindi, non dovrà installare alcun aggiornamento di sistema.

6. Risoluzione Problemi

In caso il cliente abbia bisogno di assistenza tecnica nella risoluzione di problematiche comuni, abbiamo a disposizione un servizio di chatbot capace di fornire risposte a questi problemi. In alternativa, è sempre possibile contattare Guardian Flow tramite l'e-mail di supporto, facilmente reperibile nella sezione "Contattaci".

7. Sicurezza

All'utente saranno fornite credenziali d'accesso al proprio account, all'acquisto del sistema, nella casella e-mail indicata all'acquisto, una volta effettuato l'accesso potrà, a suo piacimento, inserire una nuova password a patto che rispetti le regole di sicurezza imposte.

La password deve essere composta da almeno:

- 8 caratteri;
- una lettera maiuscola;
- una lettera minuscola;
- un numero;
- un carattere speciale.

Al momento del primo accesso, il cliente sarà automaticamente registrato come Amministratore e avrà assegnati i permessi esclusivi, consentendogli di utilizzare funzionalità privilegiate. Tra questi la possibilità di inserire Utenti Subordinati a sé o di creare altri Amministratori, inserendo i loro dati personali e le credenziali per effettuare l'accesso veranno inviate nella e-mail specificata dall'amministratore.

8. Glossario

B

Back-end
Il "back-end" è la parte di un'applicazione informatica o di un sistema software che gestisce le funzionalità non visibili direttamente agli utenti finali. Si occupa delle operazioni di elaborazione dei dati, della logica di business, dell'accesso al database e di altre attività legate al funzionamento interno di un'applicazione.

Baseline
Un insieme di dati correlati, rappresentativi del traffico ordinario e delle attività aziendali, utilizzato per addestrare un modello di intelligenza artificiale. Questo gruppo di dati è essenziale per garantire che il modello sia ben adattato alle condizioni reali dell'azienda, consentendo un apprendimento efficace e una migliore capacità predittiva nei confronti delle operazioni quotidiane dell'azienda.

I

Intelligenza artificiale
L'intelligenza artificiale (IA) è un campo dell'informatica che si occupa dello sviluppo di sistemi e programmi informatici in grado di eseguire compiti che richiedono tipicamente l'intelligenza umana. Questi compiti includono il riconoscimento di modelli, il linguaggio naturale, la risoluzione di problemi, l'apprendimento e l'adattamento a nuove situazioni.

Isolation Forest
È un algoritmo di rilevamento delle anomalie basato su alberi decisionali. La sua caratteristica distintiva è la capacità di isolare rapidamente le anomalie in un set di dati. Funziona creando alberi di decisione in cui le anomalie vengono segregate con poche iterazioni, mentre gli esempi normali richiedono più passaggi per essere separati.

F

Flessibilità
La flessibilità è la capacità di adattarsi, piegarsi o modificarsi facilmente in risposta a nuove circostanze, cambiamenti o esigenze.

Framework
Un framework è un insieme di strumenti, librerie, convenzioni e linee guida che forniscono una struttura comune per lo sviluppo di software

Front-end
Il "front-end" è la parte di un'applicazione informatica o di un sito web con cui gli utenti interagiscono direttamente. È la parte visibile e accessibile agli utenti finali, responsabile dell'aspetto grafico, dell'interfaccia utente e dell'esperienza complessiva dell'utente.

G

Gdpr

Il GDPR, o General Data Protection Regulation, è un regolamento dell'Unione Europea che disciplina la protezione dei dati personali dei cittadini dell'UE e dello Spazio Economico Europeo (SEE)

S

Scalabile
Un sistema scalabile è un sistema che può gestire un aumento del carico o delle risorse senza subire una significativa perdita di prestazioni. In altre parole, la scalabilità si riferisce alla capacità di un sistema di adattarsi e crescere efficacemente per gestire una quantità crescente di lavoro, traffico o dati.

Sistema
Un sistema informatico è un insieme di componenti hardware, software, reti e procedure che lavorano insieme per elaborare, memorizzare e trasmettere informazioni.

Software
Software indica l'insieme dei programmi e delle istruzioni che controllano il funzionamento di un sistema informatico. Può includere sistemi operativi, applicazioni e altri programmi.

Subordinati
Un account subordinato è un account collegato a un account principale, spesso con livelli di accesso o autorizzazioni inferiori.

Supervisionato
Nell'ambito del machine learning, l'apprendimento supervisionato è un tipo di approccio in cui un modello viene addestrato su un insieme di dati di input e output noti.