NTT DATA
Trusted Global Innovator

Come sfruttare le potenzialità del Machine Learning in cloud

Proteggere un dato remoto senza comprometterne l'utilizzabilità

WHITE PAPER / OMOMORFICA

Indice dei contenuti

Introduzione

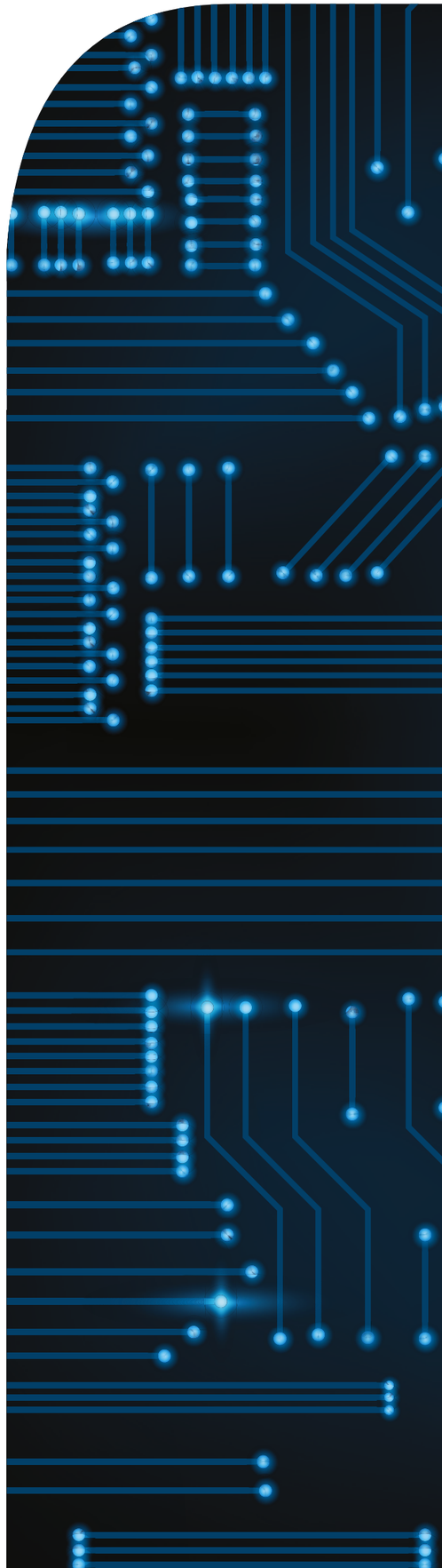
Dalle chiavette USB al Cloud Computing

Negli ultimi anni il concetto di **fruizione del dato** è andato via via evolvendosi di pari passo con l'evoluzione tecnologica, accentuando il focus sulla facilità di accesso soprattutto in mobilità.

I metodi classici di archiviazione (le memorie fisiche) non riescono più a sopperire a tali richieste e questo ha portato le grandi aziende hi-tech ad investire nel mondo del **Cloud Computing**.

In poche parole, il Cloud è una "nuvola" di dati e servizi, sempre accessibile tramite Internet da qualsiasi dispositivo ed in qualsiasi luogo.

Quello che si perde in materia di possesso del dato lo si guadagna in fruibilità e scalabilità, potendosi avvalere di hardware personalizzabili per servizi di calcolo, risorse di archiviazione, database, utilizzo di software, servizi di data intelligence...



Vantaggi e svantaggi dell’utilizzo del Cloud Computing

• Scalabilità

Non è necessario allocare in anticipo una determinata quantità di risorse ma è possibile usufruire solo di quelle di cui si ha realmente bisogno, espandendole o riducendole al variare delle proprie necessità.

• Agilità e Velocità

È possibile accedere in modo semplice a diverse tecnologie, anche molto diverse fra loro, integrandole e permettendo un notevole incremento dell’automatizzazione dei processi aziendali.

• Accessibilità

Dati e applicazioni basate sul Cloud sono accessibili praticamente da qualsiasi dispositivo collegato a Internet, da qualsiasi regione geografica ed in qualsiasi momento

• Costi

Si evitano spese dovute all’acquisto, configurazione, installazione, manutenzione e dismissione di hardware e software, in favore di una spesa variabile, pagando solo per le risorse IT realmente consumate.

• Disaster Recovery

I servizi basati sul Cloud forniscono un rapido recupero dei dati per tutti i tipi di scenari di emergenza, dai disastri naturali alle interruzioni di corrente.

• Connessione Internet

Internet è una componente necessaria per il Cloud, in assenza di connessione non è possibile accedere ai propri dati o servizi. In un mondo interamente Cloud paesi o aree geografiche prive di un buon accesso a internet rischierebbero di rimanere isolate.

• Perdita di Controllo

La gestione delle infrastrutture Cloud è interamente governata e monitorata dal fornitore di servizi, aumentando l’utilizzabilità ma riducendo sensibilmente il controllo da parte dell’utente finale.

• Data Privacy

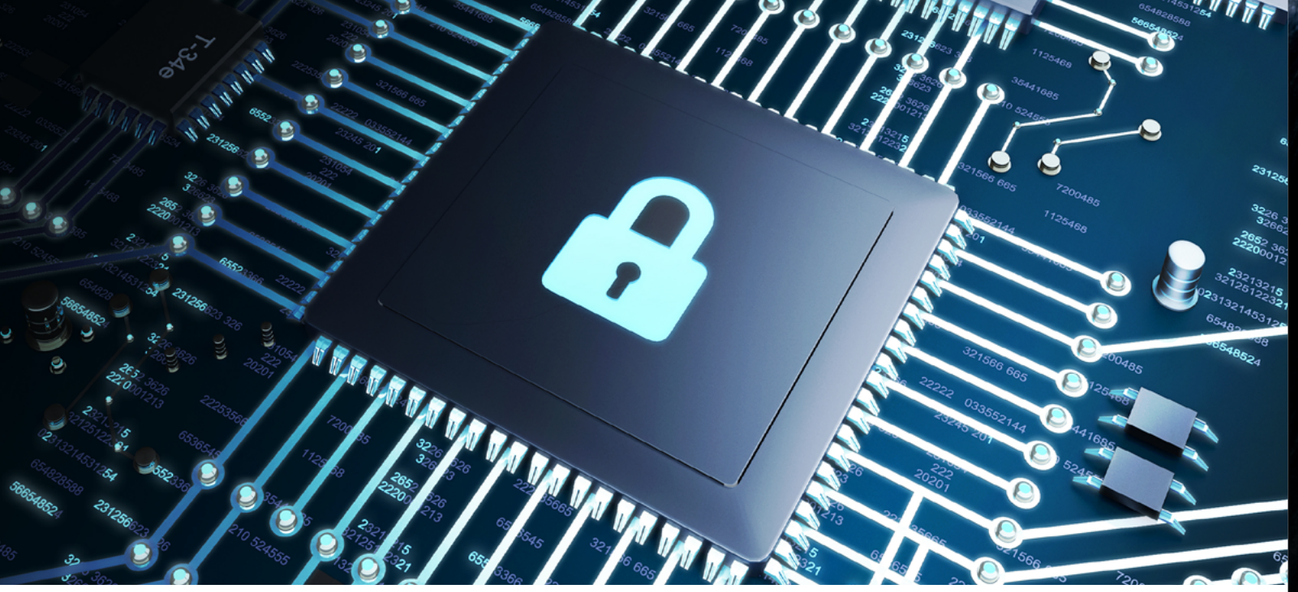
La facilità di condivisione delle risorse può portare a problemi di spionaggio industriale e perdita, manipolazione o, in generale, esposizione dei dati a terzi. Questo può condurre a violazioni della privacy e delle normative vigenti in materia di protezione del dato.

• Security

La natura pubblica e condivisa dei servizi in Cloud presta il fianco ad attacchi di tipo Data Breach o DoS/Ddos, causando disservizi che potrebbero portare ad impatti economici negativi per il cliente.

• Dipendenza dal Vendor

Anche detto “vendor lock-in”, consiste nella dipendenza diretta dall’implementazione della tecnologia di un singolo fornitore di servizi Cloud e nell’impossibilità di migrare verso un altro fornitore senza costi sostanziali, vincoli legali o incompatibilità.



È possibile ridurre il rischio di violazioni della privacy per i dati in Cloud?

Esistono molte risposte a questa domanda, sicuramente una delle più immediate è quella della **cifratura**. I crittosistemi classici garantiscono la protezione del dato ma ne compromettono l'utilizzabilità, a meno di riportarlo alla sua forma in chiaro. È proprio in questo contesto che trova spazio un'innovativa tecnica crittografica: la **crittografia omomorfica**.

La crittografia omomorfica permette di effettuare **operazioni aritmetiche su oggetti cifrati** senza doverli decifrare e quindi senza avere mai accesso ai dati di partenza.

Il risultato calcolato è in forma cifrata ma, se venisse decifrato, questo risulterebbe uguale a quello ottenuto operando con i dati in chiaro.

LA CRITTOGRAFIA OMOMORFICA, IN PRATICA

Si tratta di un argomento estremamente complesso dal punto di vista matematico ma del quale si può dare un'idea intuitiva con un semplice esempio.

5 + 10

CIFRARE CON LA CHIAVE PUBBLICA

X + YZ = PJI

DECIFRARE CON LA CHIAVE PRIVATA

15

Fra le molte applicazioni della crittografia omomorfica una preponderante è quella nell'healthcare. Il tema del trattamento dei dati personali in ambito medico, infatti, riveste un ruolo di massima attenzione nel processo di digitalizzazione del contesto sanitario. Ad esempio l'analisi predittiva è difficilmente delegabile a terze parti a causa dell'alto livello di privacy richiesto dall'utilizzo di dati medici. L'utilizzo di dati cifrati omomorficamente risolve questo impedimento inibendo l'accesso ai dati mantenendone intatta la fruibilità.

L’inarrestabile ascesa del Machine Learning

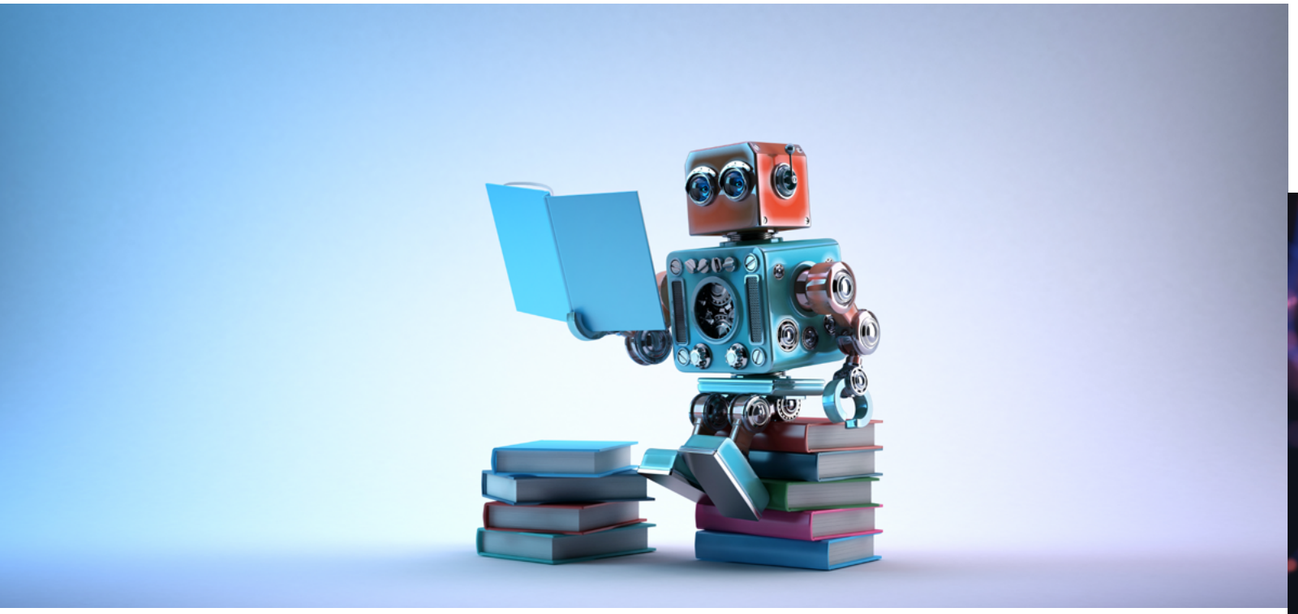
Il Machine Learning è una branca dell’**Intelligenza Artificiale** (AI) che studia differenti meccanismi che permettono ad una macchina intelligente di migliorare le proprie capacità e prestazioni nel tempo.

La macchina, quindi, sarà in grado di migliorare le proprie capacità, risposte e funzioni tramite l’esperienza. L’idea nasce dalla teoria che **i computer possano imparare ad eseguire compiti specifici** senza essere programmati per farlo, grazie al riconoscimento iterativo di schemi tra i dati.

In questo modo un computer è in grado, ad esempio, di individuare informazioni, anche sconosciute, senza che venga segnalato esplicitamente dove cercarle.

Le **applicazioni** del Machine Learning sono tanto numerose quanto eterogenee, alcune delle quali sono entrate a far parte della nostra quotidianità senza che ce ne rendessimo conto.
Alcuni dei principali esempi sono:

- **Motori di ricerca**: attraverso una o più parole chiave, questi restituiscono liste di risultati ottimizzate sulla base di sofisticati algoritmi di ricerca.
- **Filtri anti-spam**: basati su sistemi che imparano continuamente sia ad intercettare messaggi di posta elettronica sospetti o fraudolenti, sia ad agire di conseguenza.
- **Prevenzione di frodi, furti di dati ed identità**: in ambito finance gli algoritmi imparano ad agire mettendo in correlazione eventi, abitudini degli utenti, preferenze di spesa, ecc. riuscendo così ad identificare in tempo reale eventuali comportamenti anomali.
- **Medicina predittiva**: gli algoritmi imparano a fare previsioni sempre più accurate per valutare i rischi di epidemie oppure per effettuare diagnosi di tumori o malattie rare in modo accurato e tempestivo.

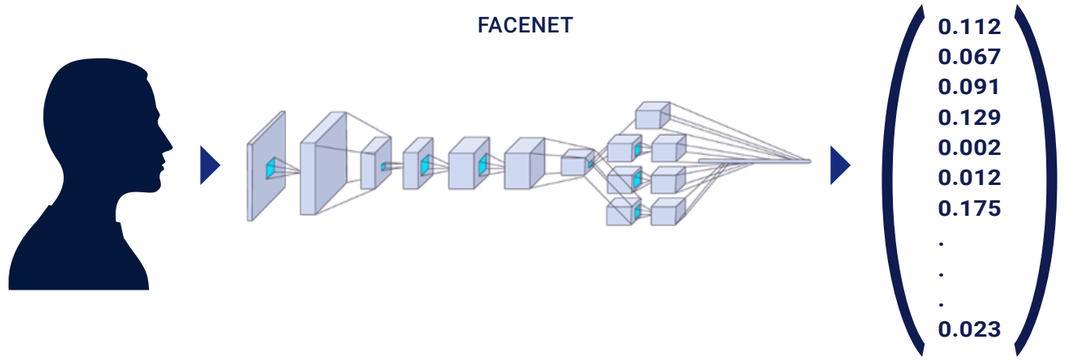


Come istruire una macchina a prendere decisioni in modo autonomo basandosi sull’esperienza?

Una delle tecniche più diffuse è quella di utilizzare le **reti neurali artificiali**.

Una rete neurale è un modello matematico composto da “neuroni” artificiali, ispirato dalla **semplificazione di una rete neurale biologica**. Possiamo considerarla come una scatola nera con diversi strati intermedi in cui l’informazione viene processata per generare il risultato finale.

Le funzioni che possono essere svolte da una rete neurale sono molteplici, ad esempio possono essere progettate per svolgere compiti di classificazione o **estrapolare pattern dall’immagine di un volto** per il riconoscimento facciale. **FaceNet** è forse l’esempio più famoso in questo contesto: sviluppata da Google, è in grado di tradurre l’immagine di un volto in un vettore numerico in maniera univoca.



Machine Learning & Cloud Computing

Tramite le reti neurali è possibile elaborare dati biometrici, come immagini del volto, impronte digitali o scansioni dell’iride, e più in generale **dati sensibili**.

Nel caso in cui la rete neurale sia on premise, la privacy dei dati viene mantenuta dalla sicurezza del singolo device.

Nel caso in cui invece la rete neurale viva su Cloud, le informazioni sensibili gestite da essa diventano vulnerabili ad attacchi. Questo problema si presenta ogni qualvolta si voglia condividere su Cloud un sistema di Machine Learning.

Come e perché trasferire un sistema di Machine Learning in ambiente Cloud

In primis per la diffusione su larga scala che il Cloud ha avuto nell'ultimo decennio, grazie a tutti i vantaggi ad esso collegati, elencati precedentemente.

Inoltre, spostando i processi di Machine Learning su Cloud, si alleggerisce il carico di lavoro del personal computer dell'utente che utilizza il servizio.
Per poter elaborare dati sensibili su Cloud tramite algoritmi di Machine Learning, senza metterne a rischio la privacy, la soluzione ottimale è ricorrere alla crittografia omomorfica.
In questo modo l'algoritmo stesso non si trova mai a gestire informazioni sensibili in chiaro, ma riesce comunque a manipolarle in maniera coerente ottenendo un output che, se decifrato dall'utente autorizzato, sia significativo.

Qualche esempio applicativo:

- **Sanità**

Sono numerose le applicazioni del Machine Learning nella ricerca, come ad esempio reti neurali per l'analisi dei tumori. Grazie alla crittografia omomorfica sarebbe possibile condividere i dati raccolti dall'ospedale con un ente esterno, mantenendo il massimo della privacy.

- **Finanza**

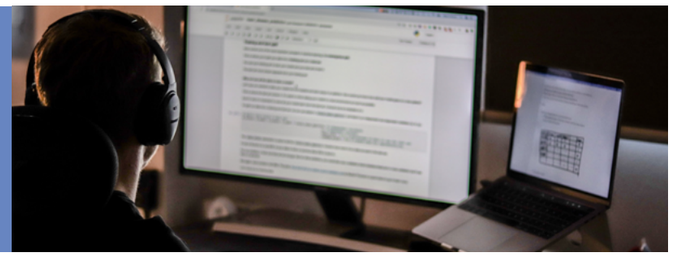
Facendo leva sulle potenzialità della crittografia omomorfica, sarebbe possibile per le aziende mettere in comune su Cloud dati sensibili degli utenti, senza rischi per la privacy, al fine di creare modelli antiriciclaggio di Machine Learning.

- **IAM**

Molti protocolli di autenticazione basati su dati biometrici implementano elementi di Machine Learning. Trasferire sul Cloud tali protocolli, sfruttando le proprietà della crittografia omomorfica per proteggere i dati, eviterebbe vulnerabilità di tipo Data Breach.

- **Smart City**

Con la crittografia omomorfica sarebbe possibile raccogliere nel Cloud ed utilizzare dati dei cittadini senza comprometterne la privacy.



Conclusioni

In conclusione, perché Cloud Computing e Machine Learning?

Perché grazie alla crittografia omomorfica è possibile trarre vantaggio da entrambe le tecnologie aprendo un ventaglio di incredibili opportunità tecnologiche e creando valore anche in contesti storicamente lontani dal mondo IT, il tutto garantendo una protezione del dato assoluta.

Key takeway

1 Gli algoritmi di Machine Learning richiedono una quantità di potenza di calcolo variabile ed elaborano una grande mole di dati. Per soddisfare queste esigenze una valida opzione è rappresentata dal Cloud Computing che, grazie alla sua scalabilità, permette di evitare spese per l'acquisto e la manutenzione di hardware e software on-premise.

2 I dati trattati in ambito Machine Learning spesso sono di tipo personale e quindi estremamente sensibili, come ad esempio informazioni biometriche, finanziarie o sanitarie; di conseguenza richiedono un alto livello di protezione.

3 La crittografia omomorfica è un tipo di crittografia che permette di eseguire calcoli su dati cifrati senza doverli prima decifrare. L'output di un'operazione omomorfica è quindi in forma cifrata e, una volta decifrato, risulterà uguale a quello che si sarebbe ottenuto operando con i dati in chiaro.

4 Il processo di Supply Chain Risk Management di NTT DATA prevede la classificazione dei fornitori e la relativa differenziazione dei controlli, in relazione ai dati gestiti dai fornitori e alla tipologia del servizio erogato.

NTT DATA
Trusted Global Innovator

Francesco Leccese
Consultant

Security Architectures & Innovation

Ilaria Ciarletti
Consultant

Information & Data Protection Practice Security

Giulia Ciccone
Junior Consultant

Information & Data Protection Practice Security

NTT DATA aiuta le organizzazioni a orientarsi nella rapida evoluzione delle tecnologie, a rispondere alle crescenti aspettative dei clienti e, attraverso l'innovazione e la profonda esperienza nel settore, mette a disposizione le competenze e le risorse per guidare lo sviluppo digitale. Offriamo consulenza in ogni fase di progetto, da una prima fase di strategia e concept, passando dagli impatti sui processi, per arrivare all'implementazione finale. Advisory, Design, Tecnologia e Operation sono solo alcune delle nostre aree di competenza. NTT DATA ha sede a Tokyo con oltre 123.000 professionisti in oltre 50 Paesi in tutto il mondo. www.nttdata.com/it

NTT DATA