Machine learning: dallo sviluppo alla validazione dei modelli *credit risk* …sei anni dopo…

Fabio Salis, Chief Risk Officer

ABI – Supervision, Risks & Profitability

Milano, 13 giugno 2024

bancodesio.it

Premessa

☞ *Convegno nazionale ABI 2018 – Roma:*

Run 2 Restart under new normality: sviluppo e validazione dei modelli *credit risk* dopo il *TRIM* e le EBA Guidelines

☞ *Convegno nazionale ABI 2024 – Milano:*

Machine learning: dallo sviluppo alla validazione dei modelli *credit risk*

… sei anni dopo…

Dove ci eravamo lasciati?

ABI 2018: Sviluppo e validazione dei modelli credit risk dopo il TRIM e le EBA Guidelines (1/2)

Nel 2018 **il quadro normativo si stava intensificando** per lo sviluppo e la Validazione dei modelli di Credit risk ed era in grande evoluzione con la pubblicazione di diverse linee guida EBA e le TRIM guide, che avrebbero apportato **grandi modifiche ai framework di sviluppo e validazione dei modelli di Credit risk**

(a) institutions shall have robust systems in place to validate the accuracy and consistency of rating systems, processes, and the estimation of all relevant risk parameters. The internal validation process shall enable the institution to assess the performance of internal rating and risk estimation systems consistently and meaningfully;

(b) institutions shall regularly compare realized default rates with estimated PDs for each grade […]. Institutions using own estimates of LGDs and conversion factors shall also perform analogous analysis for these estimates. […] The institution shall document the methods and data used in such comparisons. This analysis and documentation shall be updated at least annually; […] *(Art. 185 Reg. UE 575/2013)*

In sinergia rispetto alle disposizioni del CRR**, l'EBA** ha pubblicato una serie di **RTS** che descrivono in dettaglio **i requisiti che guidano le funzione di Sviluppo e Validazione Interna**:

- **EBA/RTS/2016/03** –**"**Final draft **RTS** on the specification of the **assessment methodology** for competent authorities regarding compliance of an institution with the requirements to use the **IRB Approach"**
- **EBA/GL/2017/16** –"Final Report: **Guidelines** on **PD estimation**, **LGD estimation** and the **treatment of defaulted Exposures"**
- **EBA/GL/2019/03** – "Final Report: **Guidelines** for the estimation of LGD appropriate for an economic downturn ('Downturn LGD estimation')"
- **EBA/GL/2020/05** – "Final Report: **Guidelines** on **credit risk mitigation** for institutions applying the **IRB approach** with own estimates of LGDs"
- **EBA/GL/2016/07** – "Final Report: **Guidelines** on the application of the **definition of default** under Art. 178 of 575/2013 EU"

- Oltre ai suddetti requisiti, la **Banca Centrale Europea** ha dettagliato, attraverso la pubblicazione della **Guida TRIM**, ulteriori requisiti per le funzioni di Sviluppo e Validazione Interna

Tenendo conto dei requisiti precedenti, la Validazione Interna:

- **dovrebbe riguardare** non solo i modelli ma **anche i processi e i sistemi IT**
- non dovrebbe limitarsi alla «***first application phase***», eseguendo **periodicamente** le proprie attività di convalida («ongoing validation»)

I **principali aspetti** che caratterizzano la Funzione di Validazione Interna sono:

- potenziale necessità di **collegare il framework di validazione** a **livello di gruppo** (sia dal lato organizzativo che metodologico)
- **panoramica** delle richieste di validazione come ad esempio **«first-validation» / «model change»** dei precedenti modelli interni

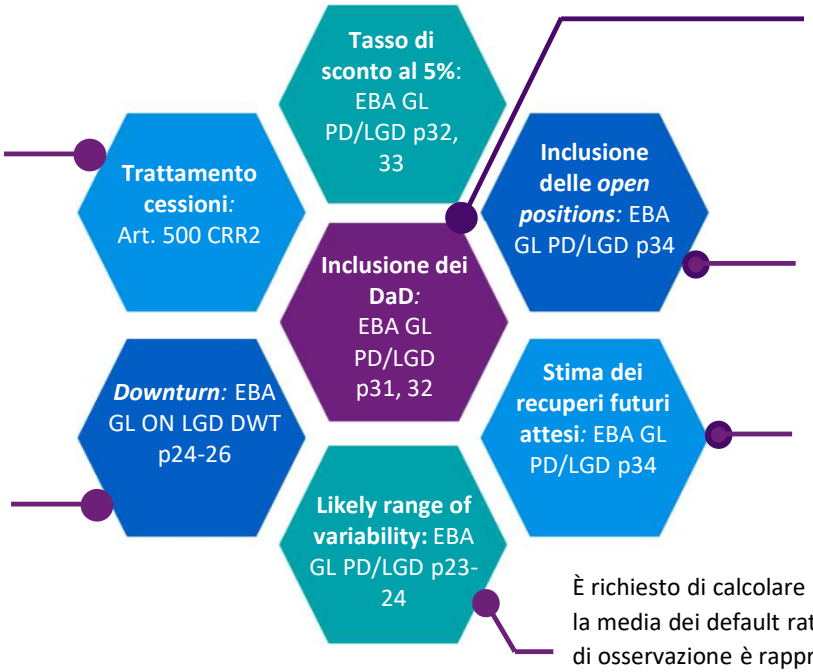
Dove ci eravamo lasciati?

ABI 2018: Sviluppo e validazione dei modelli credit risk dopo il TRIM e le EBA Guidelines (2/2)

Le novità normative avevano condotto ad una varietà di sfide in **diversi ambiti dei modelli Credit risk PD, LGD e CCF** relative a **Sviluppo e Validazione**

Per le **cessioni** avvenute **nel periodo 09-2016 – 06-2022** è previsto un ***waiver***, che permette di **mitigarne** l'**effetto negativo**, assegnando a tali posizioni la LGS delle *open positions*.

La nuova normativa richiede, ove possibile, la stima di una **LGD *downturn* basata sull'impatto** storicamente **osservato** sul parametro di una **situazione di *downturn***, identificato tramite un set di variabili macro-economiche rilevanti, tra cui GDP e UR. Il *downturn* deve corrispondere ai periodi in cui le **variabili** selezionate **assumono i valori più severi**, relativamente agli **ultimi 20 anni**, e deve avere una **durata minima** di **un anno**.



Le EBA GL richiedono di **considerare** i ***drawings after default***. In particolare, se questi sono inclusi nella stima del CCF, è richiesto che siano considerati anche nel denominatore della LGD.

È richiesto di **considerare tutti i default** nella stima della LGD, considerando quindi anche le **posizioni aperte** e cedute. Per ottenere un valore realistico di LGD, bisogna stimare i **recuperi futuri attesi** per tali posizioni.

È richiesto di calcolare il **long-run average default rate** come la media dei default rate annuali osservati se il periodo storico di osservazione è rappresentativo del **likely range of variability** dei default rate annuali, in particolare, se il periodo di osservazione storico contiene un'adeguata combinazione di anni positivi e negativi.

Sei anni dopo: che cosa è cambiato?

Quali sono le *principali normative* relative a sviluppo e validazione dei modelli di Machine Learning?

Nell'ultimo decennio, l'aumento della capacità computazionale, il consolidamento di nuove metodologie di elaborazione dei dati e la disponibilità di accesso a nuove informazioni riguardanti sia individui che organizzazioni, aiutati dal diffuso utilizzo di Internet, ha incrementato lo sviluppo e l'implementazione di tecniche di Artificial Intelligence (di seguito denominata AI) nell'attività di impresa in generale e nell'attività degli intermediari finanziari in particolare.

EBA Machine Learning for IRB Models (2023)

Il DP ha l'obiettivo di coinvolgere il settore e la comunità di vigilanza per indagare sul possibile utilizzo del ML per i modelli IRB e per costruire una comprensione comune degli aspetti generali del ML per i modelli IRB e delle relative sfide nel rispetto dei requisiti normativi.

Focus a slide 6

AI Act (2023)

L'AI Act è una proposta legislativa dell'Unione Europea (UE) che mira a regolamentare l'uso e lo sviluppo dell'intelligenza artificiale (IA) all'interno dell'UE

Focus a slide 8

PSD2 (2019)

La seconda Direttiva sui Servizi di Pagamento Europei (PSD2), ha reso le informazioni delle transazioni del cliente più accessibili e quindi più facili da utilizzare a fini di valutazione del credito.

Non si osserva ancora sul mercato un framework consolidato per la gestione e il controllo dei rischi derivanti dall'utilizzo dell'intelligenza artificiale. L'assenza di un framework non ha tuttavia frenato la diffusione dei modelli basati sull'AI a supporto di vari processi bancari (ad es. attribuzione del rating, erogazione del credito, rilevamento frodi, early warning systems, etc.).
Nella maggior parte dei casi, le funzioni di Controllo non dispongono di un framework adeguato per poter valutare l'appropriatezza di tali modelli e, pertanto, certificarne l'utilizzo

L’EBA ha pubblicato un report su Machine Learning for IRB Models

Quali sono i punti chiave del Discussion Paper sulle *tecniche di ML nei modelli IRB*?

Il report fornisce una guida completa su come le istituzioni finanziarie dovrebbero affrontare l'implementazione delle tecniche di ML nei modelli IRB, tenendo conto delle sfide complesse e delle interazioni normative

a

Le istituzioni finanziarie utilizzano le tecniche di ML principalmente per **la stima della probabilità di default (PD)** nella fase di differenziazione del rischio dello sviluppo del modello. L'uso di tecniche di ML per la validazione del modello e la valutazione delle garanzie è attualmente meno comune.

b

Nell’utilizzo del ML si deve considerare l’interazione con altri quadri normativi:

- **GDPR**: riguarda principalmente l'uso dei dati personali e le implicazioni etiche e legali, vietando l'uso di determinati dati personali per la valutazione della solvibilità.
- **AI Act**: chiarisce l'ambito di applicazione dell'AI e richiede alle istituzioni di conformarsi a requisiti aggiuntivi quando utilizzano tecniche di ML nei modelli di rischio di credito.

c

È necessario trovare un equilibrio appropriato tra le **prestazioni del modello** e la **spiegabilità dei risultati**, evitando una complessità eccessiva e garantendo una comprensione adeguata del modello da parte di tutto il board of directors. Le istituzioni devono anche assicurare che i **driver di rischio selezionati siano significativi e appropriati** e che i modelli siano adeguatamente documentati e comprensibili.

d

Le istituzioni devono prestare particolare attenzione alla **qualità dei dati**. Devono anche garantire che il personale sia adeguatamente formato per sviluppare, validare e comprendere i modelli di ML.

L’EBA ha pubblicato un report su Machine Learning for IRB Models

Quali sono i punti chiave del Discussion Paper nello *sviluppo e validazione dei modelli IRB nel ML*?

Le principali sfide nell'uso del ML includono questioni statistiche come l'**overfitting**, **problemi di competenze umane** e **difficoltà di interpretazione e spiegabilità**.

Sfide chiave nello sviluppo

- **Overfitting**: modelli ML sono molto inclini a soffrire di overfitting, pertanto è necessario porre attenzione al confronto delle prestazioni del modello misurate all'interno del campione di sviluppo con quelle ottenute utilizzando un campione di dati out of time e out of sample.
- **Competenze umane**: Lo sviluppo e la validazione dei modelli ML richiedono competenze avanzate in matematica, statistica e programmazione; è necessaria un alto grado di formazione
- **Complessità dei modelli:** data la complessità del modello, risulta difficile assicurare che i criteri di rating siano coerenti e appropriati, e che possano essere documentati in modo da permettere il giudizio umano e la conformità ai requisiti regolamentari (CRR).
- **Esigenze operative e IT**: La validazione dei modelli ML richiede risorse aggiuntive in termini di tempo, capacità computazionali e IT, nonché risorse umane. La gestione di grandi quantità di dati e l'uso di nuove fonti di dati (come i social media) aumentano le necessità operative e IT rispetto ai metodi tradizionali

Sfide chiave nella validazione

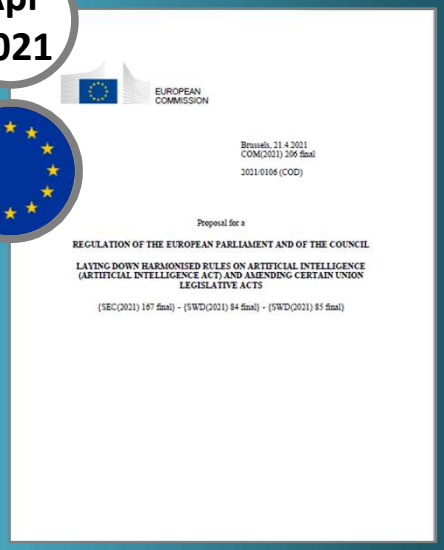
- **Interpretabilità**: trade off tra complessità, miglioramento della performance del modello e spiegabilità. Le istituzioni devono bilanciare la performance del modello con la sua interpretabilità utilizzando metriche appropriate (esempio Shapley). Ciò potrebbe rendere più difficile il rispetto del requisito della CRR di integrare lo sviluppo e l'applicazione del modello con il giudizio umano.
- **Model design**: difficoltà per la funzione di validazione che deve analizzare ed elaborare soluzione challenge sul model design. Un modello più complesso sarà più difficile da validare in modo efficace. La validazione degli iperparametri può richiedere ulteriori conoscenze statistiche.

La Commissione Europea ha iniziato a delineare le regole di fondo dell'AI ACT

Quali sono le *novità normative* introdotte dall'AI Act?

Adozione a fine 2023

Apr 2021


EUROPEAN COMMISSION

Brussels, 21.4.2021
COM(2021) 206 final
2021/0106 (COD)

Proposal for a

REGULATION OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL

LAYING DOWN HARMONISED RULES ON ARTIFICIAL INTELLIGENCE (ARTIFICIAL INTELLIGENCE ACT) AND AMENDING CERTAIN UNION LEGISLATIVE ACTS

{SEC(2021) 167 final} - {SWD(2021) 84 final} - {SWD(2021) 85 final}

A chi si Applica?

- ai **fornitori** che immettono sul mercato o mettono in servizio sistemi di AI nell'Unione;
- agli **utenti dei sistemi** di AI fisicamente presenti o stabiliti nell'Unione;
- ai **fornitori e agli utenti di sistemi** di AI fisicamente presenti o stabiliti in un paese terzo, laddove l'output prodotto dal sistema AI utilizzato nell'Unione;
- agli **importatori e ai distributori** di sistemi di AI;
- ai **fabbricanti di prodotti** che immettono sul mercato o mettono in servizio un sistema di AI insieme al loro prodotto e con il loro nome o marchio;
- ai **rappresentanti autorizzati di fornitori**, stabiliti nell'Unione.

Con che Obiettivi?

- assicurare il corretto funzionamento del mercato unico creando le condizioni per **lo sviluppo e l'utilizzo di intelligenza artificiale affidabile** nell'Unione;
- assicurare che i sistemi di AI immessi sul mercato e utilizzati siano **sicuri e rispettino** la normativa vigente in materia di **diritti fondamentali e i valori dell'Unione**;
- assicurare la certezza del diritto per **facilitare gli investimenti e l'innovazione nell'AI** chiarendo quali requisiti essenziali, obblighi e procedure di conformità e rispetto dei requisiti debbano essere rispettati.

La Commissione Europea ha iniziato a delineare le regole di fondo dell'AI ACT

Quali sono i *livelli di classificazione*?

Utilizzi proibiti dell'AI
Sono vietati modelli che manipolano il comportamento umano per aggirare il libero arbitrio degli utenti.

Sistemi AI regolamentati ad alto rischio
Sono necessari controlli attenti dei modelli durante tutto il loro ciclo di vita.

Sistemi AI che richiedono trasparenza
Sono richiesti specifici obblighi di trasparenza. Necessario rendere consapevole l'utilizzatore che sta interagendo con una macchina.

Sistemi che non implicano rischi apparenti
È raccomandata l'applicazione dei requisiti previsti per i modelli a rischio più alto.

Rischio inaccettabile
Una comprovata minaccia per la sicurezza, i mezzi di sussistenza ed i diritti delle persone

Alto rischio
Una probabile possibilità di arrecare rischio alla salute ed alla privacy delle persone o di creare danni ai servizi pubblici

Rischio limitato
Una possibile influenza del pensiero dell'utilizzatore anche in assenza di una decisione

Rischio minimo
Nessuna lesione dei diritti o la sicurezza dei cittadini.

La Commissione Europea ha iniziato a delineare le regole di fondo dell'AI ACT

In che modo agisce sui *modelli ad alto rischio*?

La Commissione EU richiede che venga **istituito, attuato, documentato e mantenuto un sistema di gestione dei rischi** con l'obiettivo di **monitorare i modelli ad alto rischio** durante tutto il loro ciclo di vita

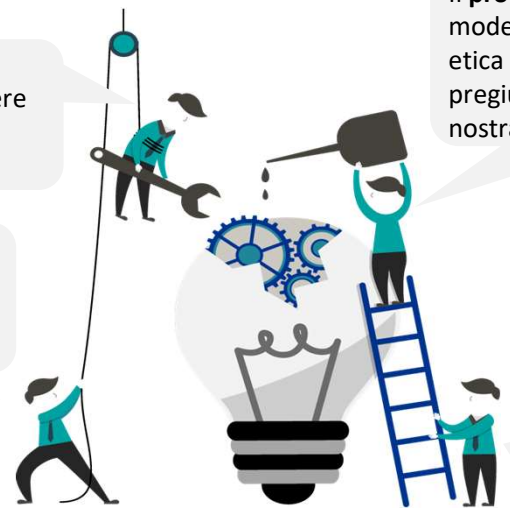
Le fasi da seguire	▪ **Identificazione** dei rischi che possono compromettere la salute, la sicurezza e i diritti fondamentali, tra tutti la **discriminazione di persone o gruppi**, ad esempio sulla base della razza, dell'origine etnica, del genere, delle disabilità, dell'età o dell'orientamento sessuale ▪ **Valutazione** di qualsiasi altro potenziale rischio ▪ Adozione di **misure** di gestione e mitigazione **dei rischi individuati**
I rischi da gestire	▪ Effettuando test per verificare la coerenza con la **finalità prevista** e la **conformità normativa** ▪ Ripetendo i test in **diversi momenti del ciclo di vita** ▪ Adottando per i test delle metriche e delle soglie **probabilistiche** che si adattano al tipo di modello identificato ed alle sue finalità
Il monitoraggio da adottare	▪ Devono essere individuati **3 set di dati con 3 finalità** diverse: Training, Validation e Test e tutti i set devono essere soggetti a controlli e verifiche ▪ Deve essere redatta una **documentazione tecnica** robusta, esaustiva e sempre aggiornata. ▪ Devono essere attivi processi di **registrazione automatica** del suo funzionamento e di **supervisione umana** ▪ Deve essere garantito che il funzionamento sia **trasparente**, **appropriato e di facile interpretazione**

È necessario un approccio robusto e industrializzato per un uso responsabile dell'AI

Quali sono i principi alla base dell'*uso responsabile dell'AI*?

Come facciamo a sapere quando l'AI inizia a prendere **decisioni al di fuori del nostro approccio**?

Rischio di approvare la **decisione sbagliata**? Le decisioni della mia AI **possono comportare sanzioni normative**?



Il **processo decisionale** del mio modello è in linea con la nostra etica e i nostri valori? È privo di pregiudizi? Può danneggiare la nostra reputazione?

Come **spiego una decisione** presa dall'AI al mio cliente, al mio responsabile o ad un regolatore in modo che la possa comprendere?
Sono **conforme** al nuovo regolamento che stabilisce norme armonizzate sull'intelligenza artificiale?

1. **Robustezza**
   Consapevolezza circa la provenienza dei dati usati per l'addestramento, le regole di addestramento, la costruzione dell'algoritmo e le logiche sottostanti le metriche prodotte, la manutenzione dall'inizio alla fine e la verifica che nessuna modifica comprometta l'obiettivo o l'intento originale del modello.
2. **Accuratezza**
   Il modello deve essere progettato con tecnologie robuste e accurate e testato attraverso delle metriche ben definite per valutarne le prestazioni.
3. **Equità**
   Il modello deve essere progettato con tecnologie robuste e accurate e testato attraverso delle metriche ben definite per valutarne le prestazioni.
4. **Privacy**
   Gli esseri umani non sono solo i principali utenti dei risultati dell'AI, ma, spesso, anche i principali produttori di input dell'AI. Questo implica che, in alcuni casi, gli esseri umani possono sentirsi esseri umani possono sentirsi a disagio quando i loro dati vengono utilizzati per l'apprendimento automatico nel contesto dell'AI
5. **Spiegabilità**
   L'umano deve essere in grado di avere contezza sul comportamento del modello, ad esempio fornendo in input attributi che non dovrebbero influenzare la scelta, osservare il risultato che deve essere coerente, soprattutto davanti ad un rischio di discriminazione.

Sfide collegate ai modelli AI

Quali sono le *principali sfide legate alla validazione*?

Un insieme coerente e definito di metriche e di parametri utili a valutare la conformità dell'intelligenza artificiale rispetto a tematiche come **sicurezza, accuratezza, correttezza, spiegabilità ed equità non esiste** ancora.

Per colmare questo divario, una proposta giunge da Babei, Giudici e Raffinetti*, che introduce parametri interconnessi, standardizzati e che hanno una radice matematica comune, **la curva di Lorenz e la sua connessione con la curva ROC** (si rimanda al paper per la definizione matematica).

La proposta prende il nome di **RGB "Rank Graduation Box".** L'uso del termine "**box**" sottolinea come il metodo RGB sia pensato per essere in continuo aggiornamento e come una scatola, possa essere riempito da nuove metriche che facciano fronte ai requisiti di conformità e gestione del rischio delle applicazioni IA che sono in continua evoluzione.

Metrica	Descrizione
RGA	Rank Graduation Accuracy, una misura di accuratezza predittiva generalista, indipendente dal tipo di modello utilizzato, dai dati utilizzati, nonché dalla natura della variabile di risposta.
RGR	Rank Graduation Robustness, una misura di **robustezza** utile a valutare la **sicurezza** associata ad un modello di mantenere il suo livello di prestazione in una varietà di circostanze avverse.
RGE	Rank Graduation Explainability, una metrica di **spiegabilità** che quantifica il rispetto di tale requisito e che consente la supervisione umana ovvero l'i intervento nel monitoraggio, controllo, mitigazione e anche cancellazione dei modelli di intelligenza artificiale, se non si comportano come previsto.
RGF	Rank Graduation Fairness, ovvero una metrica utile a stabilire il livello di **equità** di un modello AI, ovvero di verificare che l'output di tale modello non comporti una disparità di trattamento tra i diversi gruppi di popolazione, ad esempio per sesso, età e nazionalità
RGP	Rank Graduation Privacy, una metrica che quantifica la compliace dell'AI ai principi che sono definiti come privacy, ovvero tutto ciò che si riferisce in generale alle norme e alle pratiche che contribuiscono a salvaguardare l'autonomia, l'identità e la dignità delle persone.

* Cfr. «safeaipackage: a Python package for AI risk measurement»

Grazie per la cortese attenzione!

fabio.salis@bancodesio.it