

INTESA  SANPAOLO

Big data e machine learning nei nuovi modelli di rating di Intesa Sanpaolo

Dario Cavarero
Resp. Ufficio Sviluppo Modelli Retail

Abi, 22 Settembre 2020

AGENDA



PRINCIPALI INNOVAZIONI E STATO DELL'ARTE



**RIFLESSIONI FINALI SULL'UTILIZZO E L'EVOLUZIONE
DEI MODELLI DI RATING INNOVATIVI**

L'esperienza di Intesa Sanpaolo

Principali caratteristiche del nuovo modello di rating

	Modello esistente (segmenti SR)	Nuovo modello SME Retail
1 Nuova struttura del modello	<ul style="list-style-type: none"> Clienti esistenti / nuovi clienti 	<ul style="list-style-type: none"> Clienti “bancarizzati” / “non bancarizzati” (con / senza dati di conto corrente)
2 Utilizzo di nuovi dati (Interni ed esterni)	<ul style="list-style-type: none"> Modello sviluppato con dati tradizionali come i moduli di bilancio, che quindi forniscono un'immagine non attuale del cliente 	<ul style="list-style-type: none"> Basato su dati innovativi (es. C/C altre banche tramite <i>Open Banking</i>), che forniscono una panoramica sulla situazione finanziaria del cliente alla data di affidamento
3 Introduzione di algoritmi avanzati	<ul style="list-style-type: none"> Sviluppo e integrazione dei moduli basati su un approccio tradizionale (Regressione Logistica) 	<ul style="list-style-type: none"> Introduzione di algoritmi di Machine Learning per migliorare l'accuratezza del modello (relazioni non-lineari tra diversi dati)

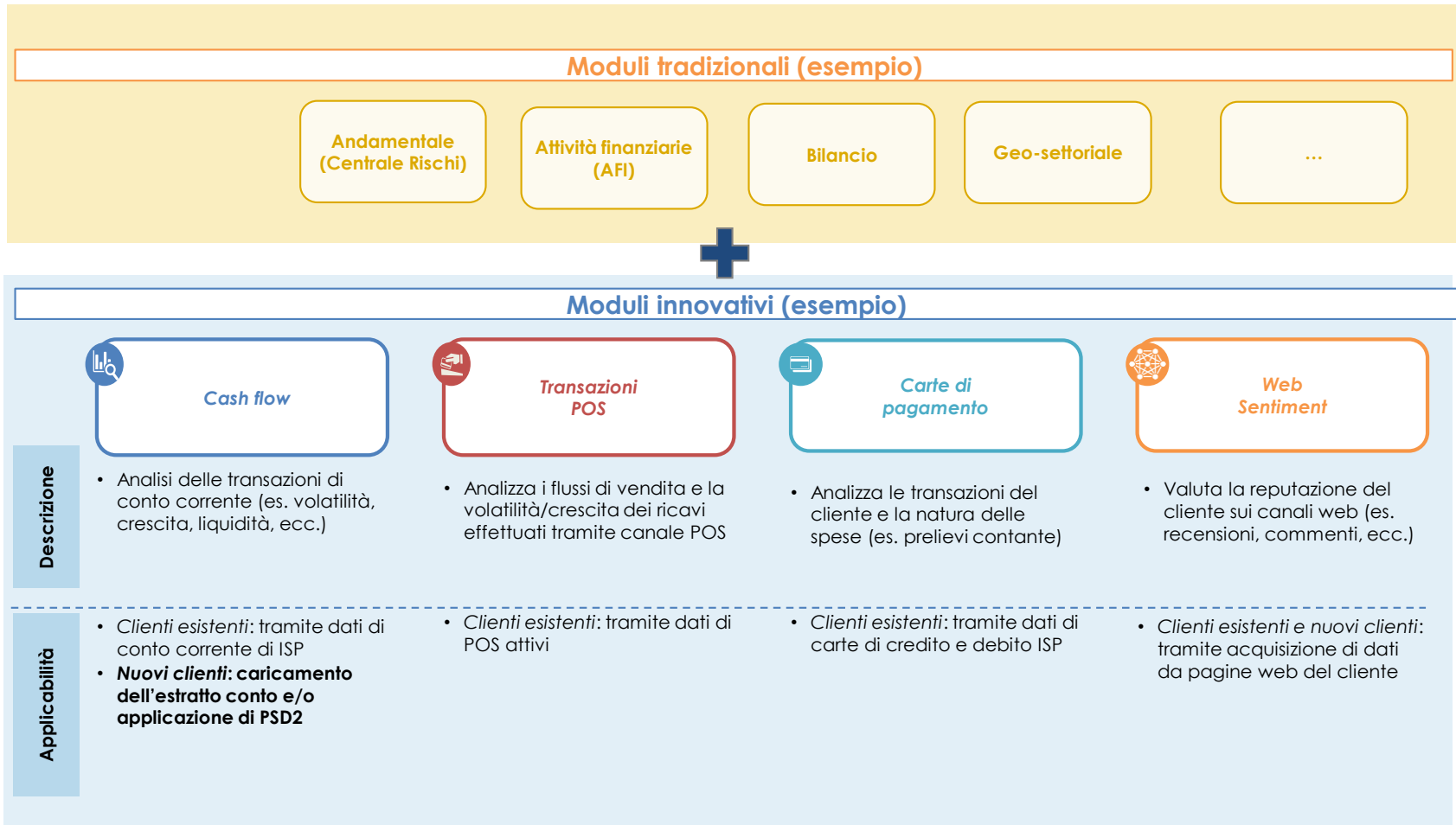
L'esperienza di Intesa Sanpaolo

I benefici più rilevanti del nuovo modello

	Modello esistente (segmenti Micro & Core)	Nuovo modello SME Retail
a Migliore accuratezza nel valutare la solvibilità del cliente	<p>Accuracy Ratio del modello (Δ)</p> <p>Clients existing Nuovi clienti</p>	<p>Accuracy Ratio del modello (Δ)</p> <p>Clients with account data available Clients without account data</p>
b Compatibilità con processi digitali	<ul style="list-style-type: none"> Richiede <u>fino a 2 settimane</u> per il calcolo rating (es. elaborazione dati, caricamento documenti) 	<ul style="list-style-type: none"> Permette un calcolo istantaneo e automatico del rating, laddove le informazioni sono disponibili

L'esperienza di Intesa Sanpaolo

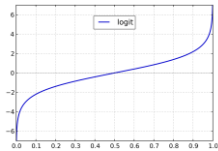
Nuove Fonti Dati



L'esperienza di Intesa Sanpaolo

Nuovi algoritmi (machine learning)

> Interpretabilità dei risultati



Logistic Regression (approccio tradizionale)

La variabile dicotomica 0/1 è spiegata tramite interpolazione logistica



Decision Trees

Un singolo albero decisionale che utilizza l'intero campione e tutte le variabili



Random Forests

Genera una serie di alberi decisionali su campioni casuali di dati e variabili e combina i risultati



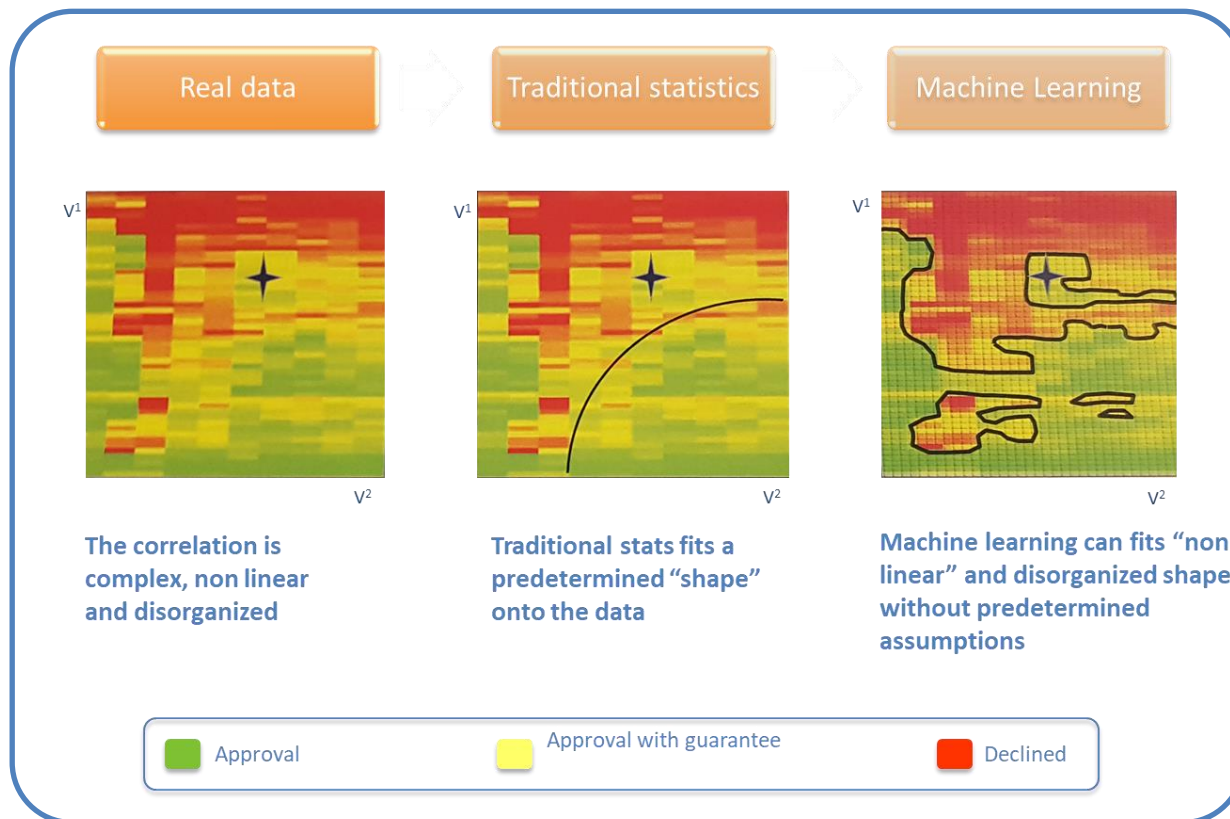
Gradient Boosted Trees (XGBoost)

Crea una serie di alberi decisionali in modo casuale sovrappesando i punti errati

> Complessità e performance

L'esperienza di Intesa Sanpaolo

Cogliere fenomeni non colti da algoritmi tradizionali e utilizzo di una grande quantità di informazioni




- Cogliere fenomeni non colti da algoritmi tradizionali che lavorano sulle correlazione e sulle medie
- Utilizzo di una grande quantità di dati (da alcuni a decine)
- Maggiore complessità

La validazione di modelli di rating con ML

Interpretabilità dei risultati

CRR 575/2013 - Art. 179 a) *“an institution's own estimates of the risk parameters PD, LGD, conversion factor and EL shall **incorporate all relevant data, information and methods.** [...] **The estimates shall be plausible and intuitive** and shall be based on the material drivers of the respective risk parameters. [...] ”*

Interpretabilità dei risultati:

- **Modello benchmark “tradizionale”**
- Partial Dependency Plot (PDP)
- Individual Conditional Expectation (ICE) plot (sensitivity al variare di un singolo ratio)
- Local Interpretabel Model-Agnostic Explanations (LIME)
-  **Shapley Additive exPlanation (SHAP)**

Si suddividono in:

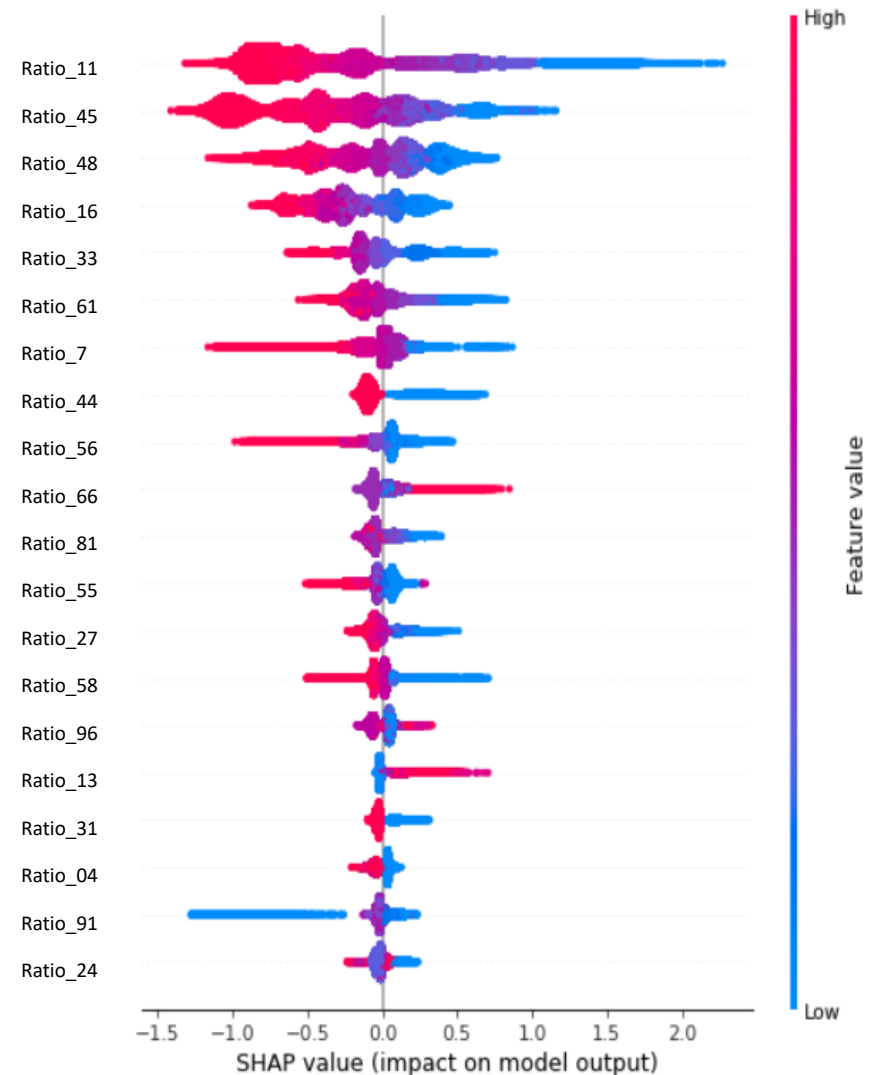
- Global Interpretability: spiega la relazione tra “inputs” iniziali e variabile target a livello complessivo
- Local Interpretability: spiega singoli cluster e singoli indicatori

Results interpretability – SHAP Summary Plot

SHAP: assegna un contributo marginale a ogni variabile considerando ogni possibile interazione con le altre variabili.

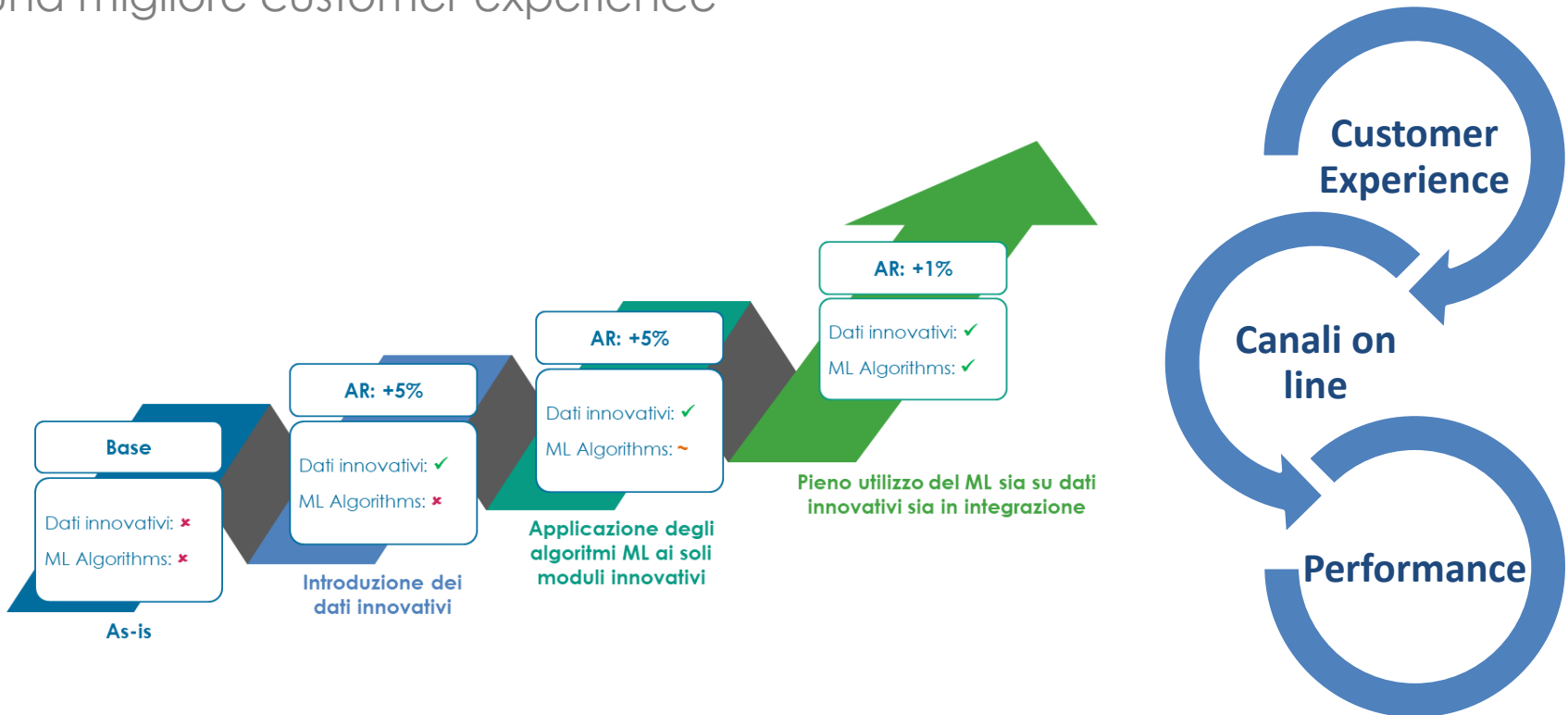
Per ogni combinazione è osservata la variazione di pd corrispondente e sulla base di essa è calcolato il peso relativo della variabile e tiene conto anche dell'incremento della performance.

Nel grafico è rappresentata il range e l'impatto di ogni variabile per ordine di importanza (feature value).



L'esperienza di Intesa Sanpaolo

Incremento di Accuracy Ratio a livello finale, ma anche in rating on line e una migliore customer experience



L'introduzione di dati innovativi ed algoritmi avanzati fornisce un miglioramento del potere predittivo rispetto ad ogni altra combinazione algoritmo (tradizionale/avanzato) – dato (tradizionale/innovativo).

AGENDA



PRINCIPALI INNOVAZIONI E STATO DELL'ARTE



**RIFLESSIONI FINALI SULL'UTILIZZO E L'EVOLUZIONE
DEI MODELLI DI RATING INNOVATIVI**

L'utilizzo del rating in banca

Regulatory and Business Strategy

Regulatory Capital

STD RWA

Large Exposure

AIRB RWA

Internal Expected

FIRB RWA

RWA Online

Stress Test & Scenario based analysis for Regulatory & Managerial Purposes

Risk Parameters definition and calibration based on forward looking scenarios

Regulatory PD, LGD, EAD

Point in Time PDs, LGDs

PD Transition Matrix



Model Laboratory

Regulatory Calculation

Managerial

Execution Type

Impairment Calculation

PD Lifetime

ECL 12M

Staging Allocation

ECL Lifetime

Stress Test & Scenario based analysis for Regulatory (TTC, Baseline, Adverse) & Managerial Purposes

Risk Adjusted Loan Pricing

Credit Monitoring

Provisioning forecasting

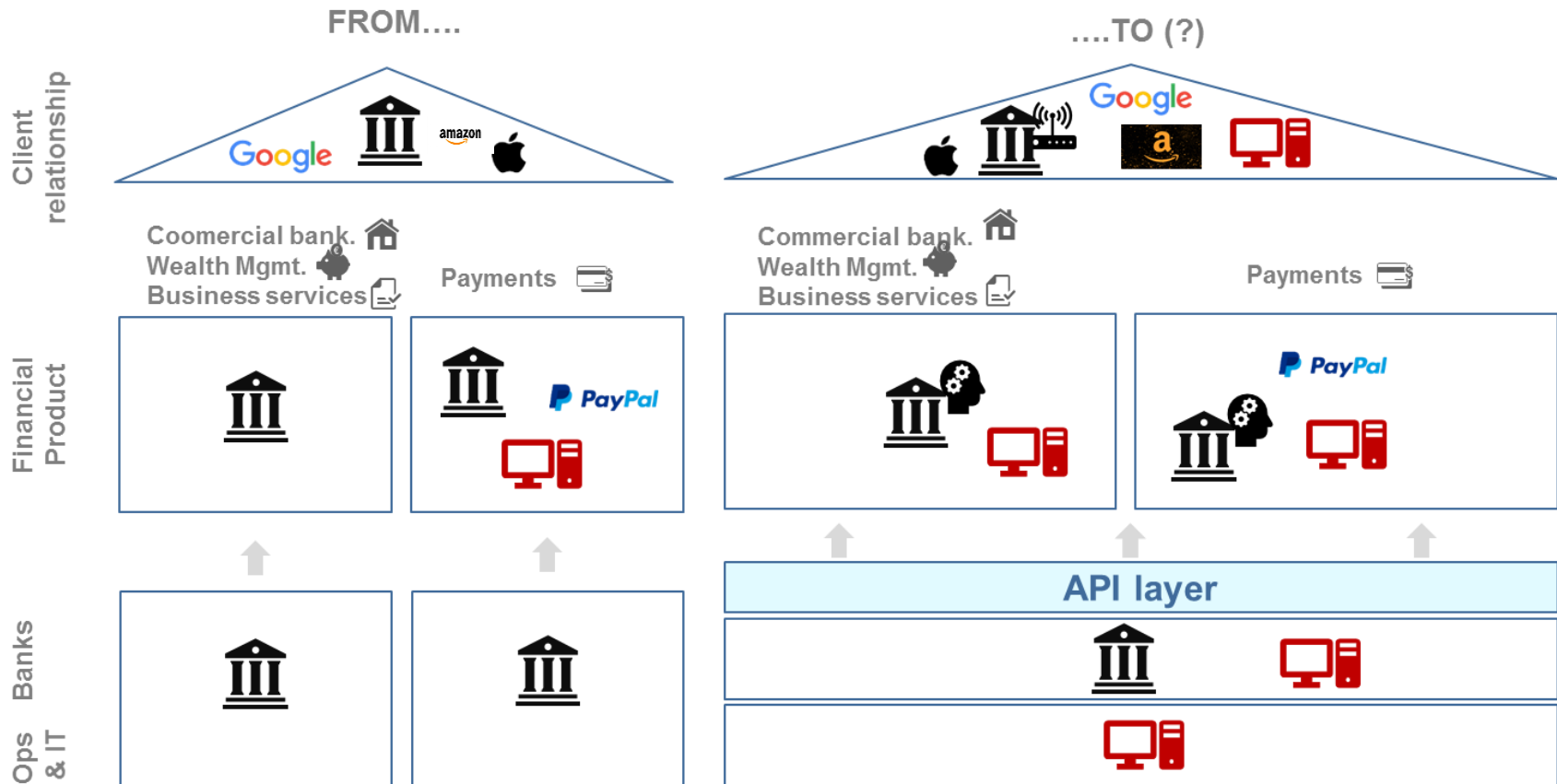
NPE & UTP



Credit Risk Strategy & Business Support

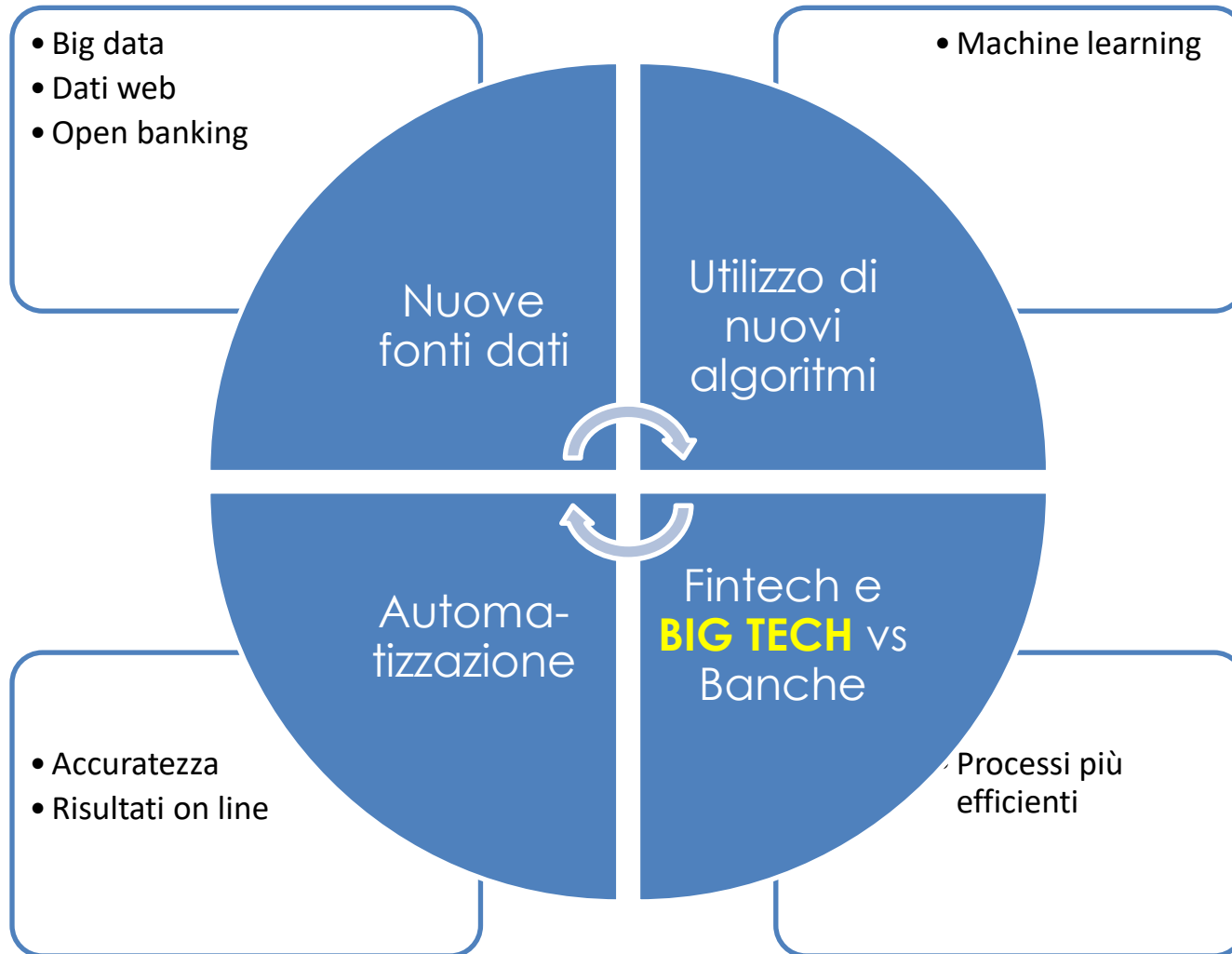
Nuovi player – la sfida tecnologica

Potenziale evoluzione a medio termine: le banche «tradizionali» e con un grado di innovazione minore potrebbero essere parzialmente disintermedate dai principali attori tecnologici. Istituzioni finanziarie intelligenti e innovative potranno confrontarsi alla pari con le Bigtech



Considerazioni Finali

Contesto di riferimento



Considerazioni Finali

La vera sfida è sui dati e sul loro utilizzo

BIS Working Papers- No 881 - Data vs collateral

by Leonardo Gambacorta, Yiping Huang, Zhenhua Li, Han Qiu and Shu Chen

*“This evidence implies that a **greater use of big tech credit** – granted on the basis of machine learning and big data – could reduce the importance of collateral in credit markets and potentially weaken the financial accelerator mechanism.”*

*“when a borrower is closely integrated into **an e-commerce platform, it may be relatively easy for a big tech to deduct the (monthly) payments** on a credit line from the borrower’s revenues that pass through its payment account .”*

*“ The **use of massive amounts of data by large technology firms (big techs) to analyse the creditworthiness of borrower firms** could replace the role of collateral in solving asymmetric information problems, with significant implications for the macroeconomy and the conduct of monetary policy. .”*

Grazie

Dario.cavarero@intesasanpaolo.com