

ANNUAL
FUNZIONI AZIENDALI DI CONTROLLO
INCERTEZZA E CONTROLLI: DA ANTITESI A SINTESI
10, 11 e 12 NOVEMBRE 2020



Laboratorio ABI KPMG: possibili soluzioni Data Analytics in ambito II e III livello

Marco Stellin
Partner, KPMG Advisory S.p.A.

ABI Associazione
Bancaria
Italiana

ABISERVIZI
ABI
FORMAZIONE

KPMG

Il laboratorio ABI KPMG

Genesi

- ✓ Gruppo di lavoro avente il principale obiettivo di valutare l'esigenza di rivedere gli accordi di servizio tipicamente adottati fra Compliance e Internal Audit

Composizione del GdL

- ✓ Partecipazione di un **panel di 15 banche**
- ✓ Partecipazione per ogni banca di una rappresentanza per l'**Internal Audit** e **Compliance**
- ✓ GdL composto da **38 soggetti**

Modalità di lavoro

- ✓ **Riunioni fisiche e virtuali**
- ✓ **Voto su piattaforma online su snodi decisionali** (*1 voto per funzione, tot 2 voti per banca*)

Obiettivi

- ✓ **Unico investimento** fra II e III livello
- ✓ **Soluzione digitale condivisa**
- ✓ **Modalità di utilizzo differenziate** per II e III livello di controllo

Overview (A) - Il metodo di classificazione funziona?



Sulla base delle loro caratteristiche personali e di operatività ci sono clienti che potenzialmente non sono classificati correttamente?

ES: clienti che operano in modo particolarmente diverso rispetto alla maggioranza degli altri clienti appartenenti alla medesima fascia di rischio

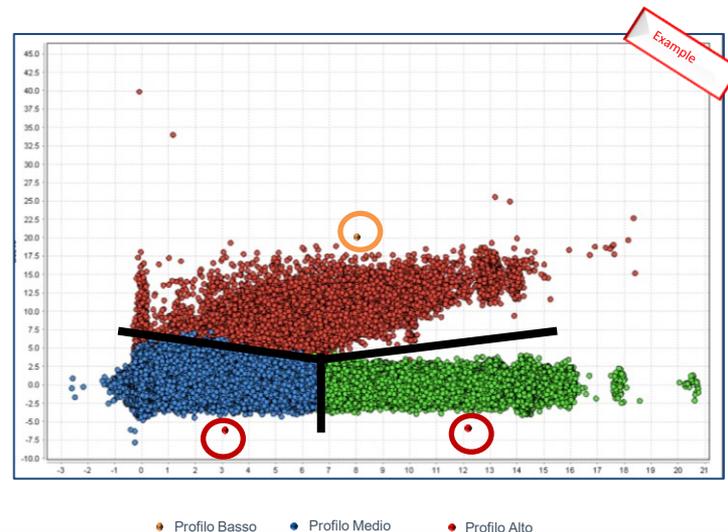


L'obiettivo è **individuare eventuali anomalie nella classificazione** della clientela, quali ad esempio:

- incoerenza tra le informazioni fornite nel questionario Mifid e le informazioni a disposizione della banca
- presenza di classi di clienti che non rispondono al questionario Mifid in maniera corretta (es: clienti di una filiale)
- riprofilazioni *ad hoc*



- *Definizione set di variabili che descrivano l'operatività della clientela (finanziamenti in essere, unexpected loss, tipologia e rischiosità degli strumenti di investimento utilizzati, VaR del portafoglio, durata media investimenti, numero di prodotti «non adeguati», ammontare in portafoglio, ecc...)*
- *Definizione set di variabili anagrafiche (professione, area di residenza, reddito dichiarato, profilo di rischio AML)*
- *Stima di un modello di clustering che consenta di raggruppare i clienti in funzione della loro operatività*
- *Produzione di un modello di anomaly detection per ogni cluster che possa mettere in evidenza i clienti che operano in modo troppo difforme rispetto al loro cluster o alla classe di rischio associata*



Overview (B) - Quali strutture commerciali si comportano in modo anomalo?



Quali sono le strutture commerciali che presentano anomalie comportamentali?

ES: Ci sono strutture commerciali che propongono solo determinate tipologie di prodotti o che effettuano tipicamente operazioni in orario di chiusura

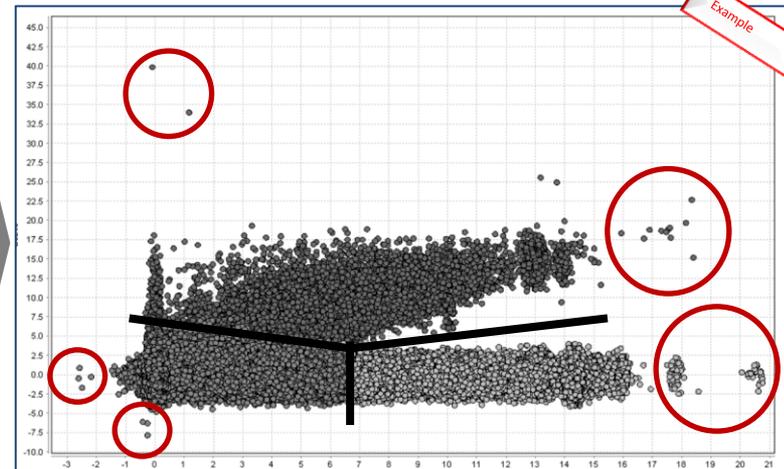


L'obiettivo è **verificare la presenza di comportamenti anomali delle strutture commerciali per cui si riscontrano valori simili di KRI al fine di:**

- Indirizzare le attività di verifica
- Individuare tempestivamente eventuali nuovi pattern di comportamenti non conformi



- *Definizione set di KRI che descrivano l'operatività delle strutture commerciali (reclami ricevuti da propri clienti, clienti con profilo alto, ordini inseriti in orario di chiusura, ordini «non appropriati», ecc..)*
- *Definizione set di variabili anagrafiche (Provincia, numero dossier gestiti, anni di attività)*
- *Stima di un modello di clustering che suddivida le strutture commerciali in gruppi simili sulla base del valore dei KRI di cui sopra. (NB: tali algoritmi tengono in considerazione sia della correlazione tra diversi KRI sia di eventuali andamenti stagionali degli stessi)*
- *Produzione di un modello di anomaly detection per ogni cluster che permetta di identificare, sfruttando precedenti verifiche, quali strutture commerciali operano in modo non conforme rispetto agli altri dello stesso cluster*



Overview (C) - Viene effettuato l'Investment churning?



Le strutture commerciali hanno effettuato operazioni non a vantaggio del cliente ma con lo scopo di aumentare le commissioni?

ES: Sono state effettuate operazioni sui dossier dei clienti che hanno generato alte commissioni ma performance troppo sotto la media?



L'obiettivo è **individuare eventuali fenomeni di investment churning al fine di:**

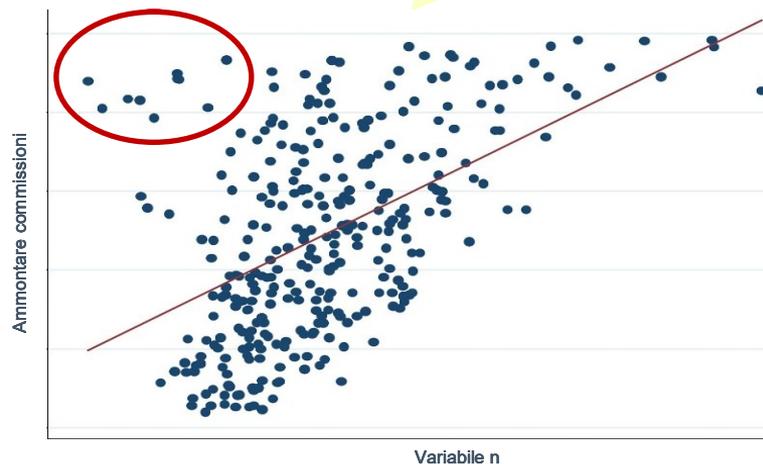
- Indirizzare le attività di verifica
- Individuare pattern di operatività anomala da parte di strutture commerciali



- *Definizione set di variabili che descrivano i prodotti e le tempistiche di investimento della clientela (Ammontare commissioni pagate, tipologia e rischiosità dei prodotti utilizzati, durata e importo medio investimenti, andamento storico del patrimonio del cliente, scostamento rispetto al portafoglio target di ogni cliente)*
- *Definizione set di variabili anagrafiche (professione, area di residenza, reddito dichiarato, profilo di rischio AML)*
- *Stima tramite una dynamic panel data regression dell'ammontare di commissioni attese per ogni sulla base delle variabili di cui sopra.*
- *Individuazione (e prioritizzazione) dei clienti che hanno pagato un ammontare probabilmente eccessivo di commissioni*



I punti che si posizionano **al di sopra** della retta di regressione rappresentano i clienti che, in media, hanno pagato più commissioni rispetto a clienti simili.



Overview (D) - Ci sono reclami che sarebbe meglio prevedere?



Quali sono i clienti che più probabilmente invieranno un reclamo alla Banca per cui verranno indennizzati?

ES: Esistono caratteristiche dei clienti o tipologie di operatività e prodotti che aumentano/diminuiscono la probabilità che determinati clienti inviino dei reclami che potrebbero essere indennizzati in modo consistente?

L'obiettivo è **calcolare la probabilità di invio di un reclamo e della relativa probabilità di successo**, al fine di:

- **Individuare nuovi presidi ex-ante** per i clienti che sulla base delle loro caratteristiche e tipologia di operatività presentano un'elevata probabilità di inviare un reclamo (es: sottoscrizione modulo aggiuntivo, limitazione all'utilizzo di determinati prodotti a maggior rischio)
- **Effettuare controlli ex-post** mirati sui clienti con una maggiore probabilità di invio di un reclamo

- *Definizione set di variabili che descrivano l'operatività della clientela (finanziamenti in essere, tipologia e rischiosità dei prodotti utilizzati, durata media investimenti, numero di prodotti «non adeguati», ammontare in portafoglio, perdita realizzata, importo minusvalenza rispetto alle perdite tollerabili dichiarate nel questionario, ecc...)*
- *Definizione set di variabili anagrafiche (professione, area di residenza, reddito dichiarato, profilo di rischio AML)*
- *Stima di un modello di supervised learning (es: logistic regression, support vector machine, random forest) che predica la probabilità e i costi che un reclamo possa andare a buon fine*
- *Identificazione delle caratteristiche della clientela e della rispettiva operatività che aumentano o diminuiscono la probabilità di invio di un reclamo*

VARIABILI	Coefficiente
Ammontare portafoglio	0,234
Profilo Alto	0,05
Profilo Medio	0,45
Profilo Basso	0,1
Durata media investimenti	0,222
Numero di finanziamenti in essere	0,07
Numero di prodotti «non adeguati»	0,53

Example

$$\begin{aligned} \text{Claim Predict Index per cliente A} &= 0,234(\text{Ammontare in portafoglio}) + 0,05(\text{Profilo Alto}) \\ &+ 0,45(\text{Profilo Medio}) + 0,1(\text{Profilo Basso}) + \\ &0,222(\text{Durata media investimenti}) + 0,07(\text{Nr} \\ &\text{finanziamenti in essere}) + \dots \end{aligned}$$

FOCUS soluzione A - Identificazione delle variabili da analizzare

Anagrafiche Connesse a dati anagrafici specifici quali: professione, area di residenza, reddito dichiarato, profilo di rischio AML.

Transazionali Connesse ad informazioni relative all'operatività della clientela quali: finanziamenti in essere, unexpected loss, tipologia e rischiosità degli strumenti di investimento utilizzati, VaR del portafoglio, durata media investimenti, numero di prodotti «non adeguati», ammontare in portafoglio



KRI attualmente in uso

Adeguatezza: incidenza di operazioni e portafogli inadeguati sul tot. complessivo.

Allineamento al portafoglio modello: allineamento degli asset del cliente rispetto al portafoglio modello.

Arretrato digitalizzazione: n. ordini dei clienti non digitalizzati.

Concentrazione di portafoglio: concentrazione operazioni svolte su titoli alto rischio e su OICR di terzi.

Concentrazione temporale: presenza di identiche operazioni compiute da più clienti del medesima struttura commerciale nella stessa giornata.

Consulenza a pagamento (evoluta): n. acquisti fuori consulenza sul totale acquisti della clientela.

Iniziativa cliente: n. di operazioni d'iniziativa cliente sul totale operazioni.

Perdite/ minusvalenze: perdite realizzate/minus maturate nel periodo.

Product Governance: discordanza tra il TM teorico negativo e TM effettivo negativo sui parametri «limitata capacità di sostenere le perdite» e «obiettivo di preservazione del capitale» per i prodotti OICR.

(Ri)profilatura: n. di clienti con profilo di rischio massimo sul tot clienti, riprofilature strumentali e portafogli incoerenti con PDR del cliente.

Rischiosità: rischio di mercato e rischio di credito del portafoglio dei clienti.



Nuovi KRI

Example

Concentrazione portafoglio: concentrazione di singolo OICR (casa e terzi) e di amministrata rischiosa/su singolo "emittente".

Profilatura clientela: presenza continuata (a due rilevazioni distanziate 12 mesi) di clienti con profilo di rischio dinamico, ma portafogli "conservativi" e privi di GP.

Iniziativa cliente OICR casa/terzi e amministrata: numero di operazioni di acquisto di iniziativa cliente su OICR della casa/ terzi/ amministrata rispetto al totale degli acquisti di OICR casa/ terzi/ amministrata (solo per clientela con consulenza a pagamento).

Product governance: incidenza di operazioni su amministrata non appropriate negoziate sui canali virtuali e aventi ad oggetto titoli a complessità elevata in "grey list"; incidenza di operazioni su amministrata disposte da clienti con basso profilo di rischio (conservativo e prudente) su prodotti con rischio alto.

Consulenza evoluta: adeguata ripartizione dei prodotti di casa e di terzi a clienti con consulenza evoluta. Previste soglie di patrimonio minime per attivare il controllo.

Adeguatezza: presidio prodotti costo di uscita; presidio Gestioni patrimoniali non adeguate.

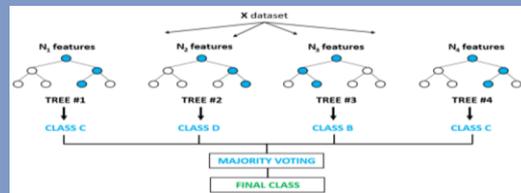
FOCUS soluzione A – Possibili algoritmi da utilizzare

Di seguito sono illustrate due tra i più efficaci algoritmi di classificazione supervisionata che potrebbero essere utilizzati per la stima del modello descritto nelle slide precedenti. La scelta degli algoritmi tuttavia dovrà essere valutata anche sulla base dei dati forniti dalla Banca.

Random Forest



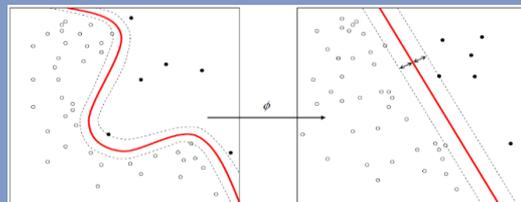
È un modello che generalizza quello degli alberi decisionali. Non necessita di alcuna assunzione sulle relazioni sottostanti al fenomeno da studiare. È praticamente immune agli effetti dell'overfitting e la presenza di outliers non influenza troppo il risultato. Produce un ranking dell'importanza delle diverse variabili. La limitazione principale di questa tecnica risulta quello di essere una black-box nella quale non è immediato risalire al perché il sistema attribuisca ad un cliente una determinata probabilità di segnalazione.



Support Vector Machine



È un modello geometrico che divide l'iperspazio delle variabili in regioni, al fine di classificare un fenomeno. Può essere utilizzato sia per predire sia variabili binarie che continue attraverso diversi metodi (i.e. epsilon regression, nu-regressionm, C-classification, nu-clasification, one-classification).



FOCUS soluzione A– Output del modello e modalità di utilizzo

Individuazione outlier

Per ogni cliente viene determinata una probabilità di appartenenza ad ognuna delle classi di profilo di rischio.

Quando la probabilità di appartenenza più alta non corrisponde al profilo di rischio assegnato al cliente (soggetti cerchiati in rosso), tale cliente deve essere attenzionato in quanto costituisce un outlier rispetto agli altri clienti presenti nella sua stessa categoria di profilo.

I campioni con cui vengono effettuati i controlli di compliance potranno essere composti esclusivamente dagli outlier secondo un approccio risk based.

La Funzione Compliance inoltre potrà valutare di apportare le dovute modifiche al corpus regolamentare interno della banca per inserire opportuni presidi che mirino a disincentivare tali comportamenti opportunistici (controlli specifici, revisione sistema sanzionatorio, revisione sistema di incentivazione, ecc.)

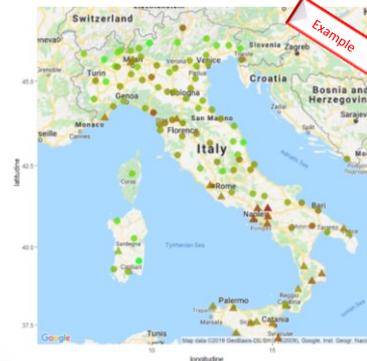
Individuazione strutture di rete a rischio

La Dashboard consente di individuare il livello di rischio delle strutture di rete sulla base del numero di clienti classificati come outlier rispetto all'attuale metodo di profilatura.

Le verifiche di internal audit potrebbero essere indirizzate verso le strutture di reti caratterizzate da maggior rischio (punti rossi nella cartina)

Cliente	Profilo di rischio attuale	Probabilità di appartenenza alle classi di profilo di rischio stimata dal modello		
		Profilo Basso	Profilo Medio	Profilo Alto
Cliente A	BASSO	90% ✓	9%	1%
Cliente B	MEDIO	5%	92% ✓	3%
Cliente C	ALTO	5%	10%	85% ✓
Cliente D	BASSO	80% ✓	10%	10%
Cliente E	MEDIO	15%	70% ✓	15%
Cliente F	MEDIO	5%	25%	70% ✗
Cliente G	ALTO	65% ✗	30%	5%

Example



Cliente caratterizzato da KRI e variabili simili a quelli dei clienti a basso rischio (es: investimenti poco rischiosi, ridotta disponibilità patrimoniale, scarsa concentrazione temporale investimenti, bassa incidenza di investimenti inadeguati sul totale)

Focus soluzione A - La Dashboard Utente

1

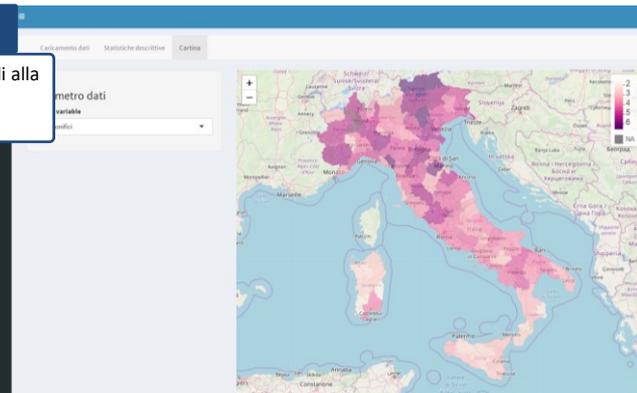
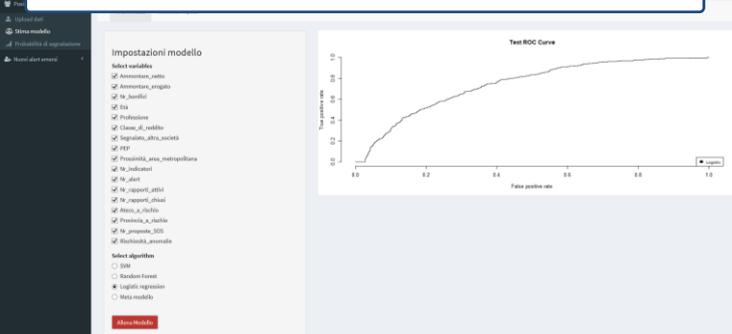
Analisi dei dati e delle variabili rilevanti ai fini Mifid

La Dashboard, consente di: i) effettuare l'upload dei dati/variabili/KRI utili alla stima del modello statistico come illustrato a slide 6; ii) svolgere analisi descrittive su tali dati (georeferenziazione e analisi grafica dei dati)

2

Stima dei modelli statistici

Una volta caricati i dati è possibile stimare il modello statistico volto all'individuazione degli outlier rispetto al metodo di profilatura della clientela ai fini Mifid adottato dalla Banca



3

Individuazione degli outlier per ogni fascia di rischio

Per ogni tipologia di profilo di rischio vengono identificati gli outlier ovvero i clienti che, sulla base delle loro caratteristiche anagrafiche e della loro operatività, sarebbe dovuta essere caratterizzata da un profilo di rischio differente.

The screenshot displays a table with columns for 'Probabilità di appartenenza', 'Classe di rischio', 'Professione', 'PEP', 'Profilo', and 'Assicurazione policy'. The table lists several risk profiles and identifies outliers. Below the table, there is a section for 'Professioni assegnate team analisti di livello medio' with a search bar and a table of assigned professions.

Probabilità di appartenenza	Classe di rischio	Professione	PEP	Profilo	Assicurazione policy
0.70	da 30.000 a 40.000	Altra	N	1	100000.00
0.70	Nuovo rischio	Comptroller	N	1	100000.00
0.75	da 30.000 a 40.000	Esploratore	N	1	100000.00
0.4	da 30.000 a 40.000	Ingegnere	N	4	40000
0.8	da 30.000 a 40.000	Ingegnere Perito	N	1	100000.00
0.70	da 30.000 a 40.000	Altra professione	N	1	100000.00
0.60	da 30.000 a 40.000	Ingegnere	N	1	100000.00

Fasi progettuali e strutture coinvolte per soluzione A

La stima del modello di *clustering* e la produzione del modello di *anomaly detection* comportano lo svolgimento delle seguenti **fasi progettuali**. Si specifica inoltre che in tutte fasi di seguito elencate è necessaria la partecipazione di c.d. Data Scientist (esterni o interni).

	 Selezione dei dati	 Soluzione tecnologica	 Selezione dei modelli	 Calcolo Outlier	 Sviluppo dashboard utente
Fasi progettuali	<ul style="list-style-type: none"> - Identificazione di una long list di variabili - Estrazione dati relativi ai clienti con data di riferimento 01/1/2018 – 31/8/2020 - Individuazione di una short list di variabili sulla base dell'effettiva disponibilità/qualità dei dati - Selezione dei clienti il cui patrimonio informativo risulta adeguato per svolgere l'analisi 	<ul style="list-style-type: none"> - Identificazione di alcuni potenziali software utilizzabili per svolgere l'analisi (es: STATA, Eviews, R, Python) - Selezione del software sulla base di: <ol style="list-style-type: none"> Open source VS Proprietary Software Costo delle licenze Compatibilità con i sistemi pre-esistenti Costi di formazione del personale e presenza materiale formativo online 	<ul style="list-style-type: none"> - Data preparation - Identificazione di alcune possibili tecniche di stima del modello di identificazione degli outlier - Stima di diversi modelli di calcolo degli outlier - Selezione dei modelli più performanti. 	<ul style="list-style-type: none"> - Individuazione degli outlier attraverso i modelli più performanti stimati in precedenza - Analisi qualitativa degli outlier individuati dal modello 	<ul style="list-style-type: none"> - Sviluppo e finalizzazione della dashboard utente che consentirà di: <ol style="list-style-type: none"> Aggiornare periodicamente la stima del modello di machine learning; aggiornare periodicamente il calcolo degli outlier; Visualizzare l'output (es.outlier) attraverso grafici e tabelle; Esportare l'output in formato excel o pdf.
Strutture coinvolte	<ul style="list-style-type: none"> - Sistemi informativi - Compliance e IA - Investimenti 	<ul style="list-style-type: none"> - Sistemi informativi 		<ul style="list-style-type: none"> - Compliance e IA - Investimenti 	<ul style="list-style-type: none"> - Sistemi informativi - Compliance e IA

Timeline progettuale per soluzione A

Di seguito viene illustrata una timeline progettuale che potrebbe essere mediamente applicabile. Si sottolinea tuttavia, che una stima puntuale delle tempistiche non può prescindere dalle caratteristiche interne ad ogni Banca (es: disponibilità di dati, algoritmi scelti per lo sviluppo del modello, applicativi in uso presso l'intermediario, ecc...) e quindi dovrà essere valutata caso per caso.

Fasi	Mese 1				Mese 2				Mese 3			
	1	2	3	4	5	6	7	8	9	10	11	12
1 – Selezione dei dati <ul style="list-style-type: none"> Identificazione di una long list di variabili Estrazione dati relativi ai clienti Individuazione di una short list di variabili Selezione clienti con adeguato patrimonio informativo per analisi 	■		■									
2 – Soluzione tecnologica <ul style="list-style-type: none"> Identificazione potenziali software di analisi Selezione software 	■	■										
3 – Selezione dei modelli <ul style="list-style-type: none"> Data preparation Identificazione tecniche di stima Stima dei modelli di clustering Selezione modelli più performanti 					■		■	■		■		
4 –Calcolo Outlier <ul style="list-style-type: none"> Produzione di un modello di anomaly detection per ogni cluster 										■		
5 – Sviluppo dashboard utente <ul style="list-style-type: none"> Sviluppo e finalizzazione dashboard utente Condivisione draft e recepimento feedback Fine tuning e rilascio definitivo 								■			■	■

Ulteriori considerazioni emerse nel corso del Workshop

In aggiunta alle diverse modalità di utilizzo del modello statistico tra la funzione Compliance ed Internal Audit così come descritte, il gruppo di lavoro sottolinea come in molte realtà bancarie:



- L'**Internal Audit** non monitori solamente la rete ma effettua anche **verifiche di processo** (in questo senso potrebbe utilizzare il modello in modo simile alla funzione Compliance)
- La **funzione Compliance** verifica anche il **rischio di «misconduct»** (in questo senso potrebbe utilizzare il modello in modo simile alla funzione Internal Audit)



Molte funzioni Compliance e Internal Audit attualmente non dispongono di propri «data scientist» per lo sviluppo ed il mantenimento di soluzioni di machine learning. A tendere tuttavia dovranno dotarsi di **specialisti funzionali con competenze statistiche** che dovranno svolgere il compito di «link» tra i data scientist (che saranno probabilmente incardinati nelle funzioni ICT) e gli addetti specializzati in Compliance ed Internal Audit.



Le tempistiche di sviluppo previste prevedono lo sviluppo di una Dashboard Utente per l'utilizzo degli esiti del modello da parte degli addetti delle funzioni Compliance e Internal Audit. Tali tempistiche **potrebbero dover essere ampliate** in caso di:

- Presenza di dati non strutturati o di bassa qualità all'interno dei database aziendali
- Sviluppo di ulteriori funzionalità finalizzate ad: i) acquisire in modo automatico i dati dai sistemi aziendali; ii) integrare gli algoritmi del modello all'interno degli applicativi in uso presso la Banca (in sostituzione della Dashboard Utente)

