

***INFORMATIVA SULLA LOGICA SOTTOSTANTE I MODELLI DI ANALISI DEL RISCHIO BASATI  
SUI DATI DELL'ARCHIVIO DEI RAPPORTI FINANZIARI***

<b>1</b>	<b>Premessa .....</b>	<b>2</b>
<b>2</b>	<b>Definizione di analisi del rischio fiscale .....</b>	<b>4</b>
<b>2.1</b>	<b>Contesto generale.....</b>	<b>4</b>
<b>2.2</b>	<b>Descrizione del processo di analisi .....</b>	<b>5</b>
<b>3</b>	<b>Basi informative utilizzate .....</b>	<b>7</b>
<b>4</b>	<b>Modelli di analisi.....</b>	<b>7</b>
<b>4.1</b>	<b>Approccio deterministico .....</b>	<b>10</b>
<b>4.2</b>	<b>Approccio stocastico .....</b>	<b>11</b>
<b>4.3</b>	<b>Algoritmo di ottimizzazione del processo di selezione delle posizioni da sottoporre a controllo .....</b>	<b>13</b>
	<b>ALLEGATO TECNICO.....</b>	<b>16</b>

## 1 Premessa

L'archivio dei rapporti finanziari costituisce un'apposita sezione dell'Anagrafe tributaria ed è una base dati che contiene le informazioni relative:

- ai conti correnti e agli altri rapporti finanziari di cui un contribuente è titolare o può disporre sulla base di deleghe o procure ad operare (c.d. "sezione anagrafica");
- alle movimentazioni contabili in forma aggregata, al saldo iniziale, a quello finale e, per alcune tipologie di conto, al valore medio di giacenza, che interessano in un anno solare ciascun rapporto continuativo, nonché alle operazioni c.d. "extra-conto", vale a dire effettuate al di fuori di un rapporto continuativo con l'intermediario finanziario (c.d. "sezione contabile").

Detto archivio è stato istituito ad opera dell'articolo 7, sesto comma, del decreto del Presidente della Repubblica 29 settembre 1973, n. 605 (sezione anagrafica) e dell'articolo 11, comma 2, del decreto-legge 6 dicembre 2011, n. 201 (sezione contabile). Il comma 4 del citato articolo 11, ha, inoltre, previsto la possibilità per l'Agenzia delle entrate di utilizzare le informazioni in esso presenti per le analisi del rischio di evasione.

Successivamente, il comma 682 dell'articolo 1 della legge 27 dicembre 2019, n. 160, ha disposto che l'Agenzia delle entrate, per le summenzionate analisi del rischio, possa avvalersi, anche previa pseudonimizzazione dei dati personali, delle tecnologie, delle elaborazioni e delle interconnessioni con le altre banche dati di cui dispone, allo scopo di individuare i criteri di rischio utili a far emergere le posizioni da sottoporre a controllo o nei cui confronti avviare attività di stimolo dell'adempimento spontaneo.

Inoltre, il comma 684 del citato articolo 1 ha statuito, nel rispetto del principio di responsabilizzazione di cui all'articolo 35 del regolamento (UE) 2016/679, che l'Agenzia delle entrate redigesse una valutazione unitaria di impatto sulla protezione dei dati (*Data Protection Impact Assessment - DPIA*) da sottoporre al parere preventivo del Garante per la protezione dei dati personali.

Con il provvedimento n. 276 del 30 luglio 2022 il Garante per la protezione dei dati personali ha espresso parere favorevole rispetto alla summenzionata valutazione di impatto privacy – uno stralcio della quale è consultabile sul sito dell’Agenzia delle entrate.

Il predetto parere contiene alcune prescrizioni che devono essere adempiute prima dell’avvio delle attività di analisi del rischio e, in particolare, la pubblicazione sul sito istituzionale dell’Agenzia delle entrate di un documento illustrativo della logica degli algoritmi, redatto conformemente alle «*Linee guida sul processo decisionale automatizzato relativo alle persone fisiche e sulla profilazione ai fini del regolamento 2016/679*» ([ARTICLE29 - Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679 \(wp251rev.01\) \(europa.eu\)](#)).

Tanto premesso, il presente documento riporta gli elementi informativi concernenti le metodologie utilizzate nell’ambito delle attività di analisi del rischio basate sull’utilizzo dei dati dell’Archivio dei rapporti finanziari. In particolare, in conformità alle citate linee guida, viene descritta la logica utilizzata dagli algoritmi, nonché le banche dati usate nelle attività di analisi del rischio.

Per salvaguardare le esigenze di riservatezza delle attività istituzionali dell’Agenzia delle entrate ed evitare l’adozione di condotte suscettibili di eludere i controlli, non si darà conto - sempre in aderenza alle predette linee guida - degli elementi di dettaglio sui singoli percorsi di indagine né delle informazioni sull’architettura informatica utilizzata.

Infine, prima di passare alla disamina delle metodologie utilizzate, saranno illustrati alcuni concetti utili per una maggiore comprensione del documento.

Per quanto concerne, infine, i diritti dei contribuenti i cui dati sono oggetto di trattamento, si rinvia all’apposita informativa - disponibile sul sito dell’Agenzia delle Entrate - nonché al D.M. 28 giugno 2022 (<https://www.gazzettaufficiale.it/eli/id/2022/07/01/22A03862/sg>) che ha disciplinato le limitazioni applicabili ai diritti degli interessati e le relative tutele.

## 2 Definizione di analisi del rischio fiscale

### 2.1 Contesto generale

In termini generali, l'analisi del rischio fiscale ricomprende le tecniche, le procedure e gli strumenti informatici utilizzati per individuare i contribuenti che presentano un elevato rischio fiscale, inteso come il rischio di operare, o aver operato, in violazione di norme di natura tributaria ovvero in contrasto con i principi o con le finalità dell'ordinamento tributario; una volta individuate le posizioni fiscalmente rischiose, le stesse sono trasmesse alle articolazioni organizzative che si occupano dei controlli, che effettuano ulteriori approfondimenti e valutazioni al fine di individuare i soggetti nei cui confronti avviare un'attività istruttoria.

In termini organizzativi, l'attività di analisi del rischio si inserisce nell'ambito più generale della gestione del rischio di non conformità (*Compliance Risk Management*), che da più di un decennio è al centro dell'attenzione dell'OCSE e della Commissione Europea che hanno fornito indirizzi e condiviso strategie per un'efficace azione *di risk management*.

In linea di principio, tale azione risponde da un lato all'esigenza di allocare al meglio le risorse delle Amministrazioni fiscali – limitate rispetto alla platea dei contribuenti – al fine di raggiungere un livello di *compliance* fiscale ottimale, dall'altro ad evitare l'attivazione di controlli nei confronti di soggetti che hanno adottato comportamenti conformi alle norme tributarie.

Si tratta, quindi, di un processo sistematico volto a stimolare l'adempimento spontaneo, ovvero prevenire e contrastare l'inadempimento. Tale processo risulta tanto più efficace quanto maggiore è la conoscenza del comportamento fiscale dei contribuenti, ricostruito attraverso lo studio approfondito e sistematico delle informazioni disponibili ed avvalendosi di strumenti di analisi avanzata dei dati.

Infatti, il patrimonio informativo di cui dispone l'Amministrazione finanziaria, per essere correttamente valorizzato ed efficacemente utilizzato, richiede strumenti e tecniche di analisi sempre

più evoluti e al passo con gli sviluppi tecnologici e metodologici, così da riuscire a trasformare i dati in informazioni e le informazioni in conoscenza, perseguendo altresì obiettivi di semplificazione, efficientamento del processo e riduzione dei costi complessivi e degli oneri a carico della collettività.

In tale contesto, l'analisi del rischio in materia fiscale, grazie alla disponibilità e al tempestivo utilizzo delle informazioni rilevanti, consente di operare interventi idonei a privilegiare la prevenzione *ex ante*, rispetto alla repressione *ex post*, nonché l'attuazione di azioni mirate e un uso più efficiente delle risorse dell'Amministrazione, circoscrivendo i controlli nei confronti di soggetti a più alto rischio fiscale e con minore impatto su cittadini e imprese anche in termini di oneri amministrativi.

Nell'ambito dei controlli fiscali, infatti, il maggiore rischio per i diritti dei contribuenti non è tanto la quantità di dati a disposizione dell'amministrazione finanziaria, ma eventualmente la loro incompletezza e la visione "parziale" che ne deriva, con conseguente possibile alterazione del quadro istruttorio complessivo e compromissione del principio di capacità contributiva e, in definitiva, dell'equità sociale e dell'equilibrio delle finanze pubbliche.

## **2.2 Descrizione del processo di analisi**

L'analisi del rischio fiscale rientra in un processo organizzativo che, tramite modelli e tecniche di analisi deterministiche e stocastiche, utilizza le informazioni presenti nelle basi dati per associare la probabilità di accadimento ad un determinato rischio fiscale, effettuando, ove possibile, anche una previsione sulle conseguenze che possono generarsi dal suo inveramento.

Detto processo può essere scomposto nelle seguenti fasi (per maggiori approfondimenti si rimanda al documento di DPIA):

1. individuazione della platea di riferimento;
2. scelta delle basi dati;
3. messa a disposizione delle basi dati;

4. analisi della qualità;
5. definizione del criterio di rischio;
6. scelta del modello di analisi;
7. verifica della corretta applicazione del modello e del criterio di rischio;
8. estrazione e identificazione dei soggetti;
9. *test* su un campione della sotto-platea di riferimento;
10. predisposizione delle liste selettive.

Nello svolgimento del processo di analisi viene sempre garantito l'intervento umano e, di conseguenza, non si fa uso di alcun tipo di processo decisionale completamente automatizzato.

Infatti, tutte le fasi summenzionate sono curate da personale specificamente individuato e, inoltre, gli esiti delle analisi non vengono utilizzati per l'emissione diretta di provvedimenti impositivi, bensì per segnalare alle strutture di controllo i contribuenti che presentano un elevato livello di rischio.

A loro volta, le strutture di controllo possono autonomamente, in base alle risultanze di ulteriori approfondimenti, decidere quali sono le posizioni rispetto alle quali attivare un'istruttoria.

L'impostazione così adottata, pertanto, consente di garantire che, in esito alle attività di analisi, non si crei alcun automatismo lesivo della sfera giuridica dei contribuenti.

Inoltre, nel momento in cui vengono utilizzate le informazioni dell'Archivio dei rapporti finanziari per l'individuazione dei rischi fiscali, i dati personali dei contribuenti vengono pseudonimizzati (cioè sostituiti con codici fittizi), in modo che, nel corso del trattamento di tali dati, non sia mai consentita la possibilità di associare i dati finanziari ad uno specifico individuo, prima che sia stata verificata la presenza di un rischio fiscale.

### **3 Basi informative utilizzate**

La scelta delle basi dati, in ottemperanza al “principio di minimizzazione dei dati” previsto dall’articolo 5 par. 1, lett. c), del Regolamento UE 2016/679 (GDPR), è effettuata limitando l’utilizzo delle informazioni a quelle maggiormente rilevanti ai fini del fenomeno tributario oggetto di analisi e della individuazione della platea di riferimento.

Le informazioni che di volta in volta vengono interconnesse con l’Archivio possono riguardare, a titolo esemplificativo, i dati dichiarativi, gli atti del Registro, i dati della fatturazione elettronica e dell’invio telematico dei corrispettivi, in funzione dello specifico contesto d’analisi.

Sono in ogni caso esclusi i dati particolari di cui all’articolo 9<sup>1</sup> e quelli giudiziari di cui all’art. 10<sup>2</sup> del Regolamento (UE) 2016/679 (d’ora in avanti, Regolamento), così come i dati che potrebbero consentire di desumere indirettamente informazioni “sensibili” (es. Stato di nascita, ammontare delle spese sanitarie e simili).

Per un maggior livello di dettaglio sulla natura dei dati utilizzati si rinvia al documento di DPIA.

### **4 Modelli di analisi**

La scelta del modello di analisi tiene conto del fenomeno fiscale che si intende indagare, del patrimonio informativo disponibile nelle basi dati e del livello di qualità dei dati, preventivamente valutato.

Le analisi del rischio di evasione possono essere basate:

---

<sup>1</sup> «..dati personali che rivelino l'origine razziale o etnica, le opinioni politiche, le convinzioni religiose o filosofiche, o l'appartenenza sindacale, nonché trattare dati genetici, dati biometrici intesi a identificare in modo univoco una persona fisica, dati relativi alla salute o alla vita sessuale o all'orientamento sessuale della persona...».

<sup>2</sup> «..dati personali relativi alle condanne penali e ai reati o a connesse misure di sicurezza...».

- sui dati dell'Archivio che, dopo l'effettuazione di una prima elaborazione eseguita in applicazione del criterio di rischio individuato, vengono, successivamente, collegati alle informazioni presenti nelle altre banche dati a disposizione dell'Agenzia delle entrate;

ESEMPIO: si individua inizialmente una platea di contribuenti che, dalle informazioni dei saldi dei rapporti finanziari, presentano elevati incrementi patrimoniali in un determinato periodo d'imposta, rispetto alla media degli incrementi registrati per le annualità precedenti; tale platea viene successivamente analizzata in base ai dati dichiarativi e ad altre informazioni rilevanti disponibili in Anagrafe tributaria, al fine di individuare eventuali casi anomali da sottoporre a controllo o, viceversa, per escludere posizioni che risultano coerenti con gli elementi informativi acquisiti

- sulle informazioni presenti nelle altre banche dati a disposizione dell'Agenzia delle entrate, alle quali, dopo l'effettuazione di una prima elaborazione eseguita in applicazione del criterio di rischio individuato, vengono successivamente collegati i dati dell'Archivio;

ESEMPIO: la platea iniziale può essere costituita dai contribuenti, titolari di P.IVA, che operano in uno specifico settore economico e che dichiarano ricavi o compensi inverosimili per la categoria di appartenenza, alla luce di alcuni elementi desumibili dalle banche dati a disposizione dell'Agenzia, quali i costi per le materie prime, le spese per il personale, etc.; il riscontro con le movimentazioni in avere risultanti sui conti correnti, indicative della possibile presenza di ricavi o compensi non dichiarati, permette di selezionare contribuenti a maggior rischio fiscale, da sottoporre ad attività istruttoria

- sui dati dell'Archivio integrati con le informazioni presenti nelle altre banche dati a disposizione dell'Agenzia delle entrate, effettuando, successivamente, le elaborazioni necessarie a riscontrare il criterio di rischio individuato;



ESEMPIO: il criterio di rischio può consistere, con riferimento ad una specifica platea di professionisti e/o titolari di attività di impresa, nello scostamento significativo tra l'ammontare dei compensi o ricavi dichiarati e gli importi dei flussi finanziari attivi risultanti dai dati dell'Archivio

- sui soli dati dell'Archivio;

ESEMPIO: il patrimonio informativo dell'Archivio può essere utilizzato per individuare, anche sulla base di appositi analisi statistiche, anomalie rilevanti nella distribuzione di alcune variabili quali ad esempio, il numero di accessi alle cassette di sicurezza, la frequenza dell'apertura e/o chiusura di rapporti, l'elevata numerosità di conti correnti e altre tipologie di rapporti finanziari.

L'analisi del rischio basata sull'utilizzo dei dati contenuti nell'Archivio può avvalersi di approcci metodologici di tipo deterministico o stocastico, che possono essere così definiti:

- **analisi deterministica:** insieme dei modelli e delle tecniche di analisi basati sul raffronto e sull'elaborazione di dati, riferiti ad uno o più contribuenti ovvero ad uno o più periodi di imposta, volti a verificare, tramite criteri selettivi fondati su relazioni non probabilistiche, l'avveramento di un rischio fiscale, in tutto o in parte definibile prima dell'avvio dell'analisi;
- **analisi probabilistica:** insieme dei modelli e delle tecniche di analisi che, sfruttando soluzioni di intelligenza artificiale ovvero di statistica inferenziale, consentono di isolare rischi fiscali, anche non noti a priori, che, una volta individuati, possono essere utilizzati per l'elaborazione di autonomi criteri selettivi, ovvero permettono di attribuire una determinata probabilità di accadimento ad un rischio fiscale noto;

Come sarà meglio chiarito nel prosieguo, l'analisi probabilistica viene utilizzata solo per corroborare gli esiti delle analisi deterministiche e, pertanto, nessun contribuente potrà essere individuato come fiscalmente rischioso esclusivamente in base alle risultanze di un modello predittivo.

#### **4.1 Approccio deterministico**

I modelli deterministici sono basati sulla verifica di relazioni non probabilistiche tra grandezze fiscali, economiche e finanziarie, al fine fare emergere possibili violazioni.

Si parla di relazioni non probabilistiche poiché i dati comparati sono connessi da una relazione di tipo certo, come può essere quella che intercorre tra gli importi indicati nelle certificazioni rilasciate dai sostituti di imposta e l'ammontare del reddito imponibile dichiarato dal percipiente.

ESEMPIO: con riferimento ad una platea di contribuenti titolari di P.IVA, la logica sottostante un modello di analisi deterministico può consistere nel confrontare gli importi delle operazioni imponibili risultanti dalle fatture elettroniche emesse in un determinato anno d'imposta con i ricavi e/o il volume d'affari indicato nei modelli dichiarativi per il medesimo intervallo temporale.

L'eventuale significativo scostamento tra dette grandezze può essere ritenuto sintomatico della presenza di ricavi sottratti a tassazione, laddove non siano ravvisabili elementi e circostanze che possano giustificare tale discrasia.

Il criterio deterministico consiste, in definitiva, nel selezionare quei contribuenti che presentano incongruenze numeriche sintomatiche di un non corretto assolvimento degli obblighi fiscali.

In quest'ottica, l'impiego dell'Archivio dei rapporti finanziari, integrato con le altre banche dati nella disponibilità dell'Agenzia delle entrate, migliora l'efficacia dei modelli deterministici, in quanto attraverso i dati finanziari può essere meglio valutata la situazione reddituale e patrimoniale del contribuente, e in ultima istanza, la sua capacità contributiva.

L'adozione di tale approccio integrato permette, tra l'altro, di concentrare le attività di controllo sulle posizioni che presentano un maggiore livello di rischio fiscale, espungendo, già in fase di analisi, quelle per le quali l'incongruenza tra redditi dichiarati e spese/investimenti effettuati sarebbe agevolmente giustificabile in sede di contraddittorio (il contribuente potrebbe eccepire, ad esempio, il ricorso a risparmi accumulati nei periodi di imposta precedenti).

Al tempo stesso, l'utilizzo dei dati finanziari disponibili nell'Archivio consente di focalizzare l'attenzione anche su profili apparentemente virtuosi in quanto caratterizzati da volumi di spesa coerenti con i redditi dichiarati, e che tuttavia presentano incrementi patrimoniali non compatibili con la capacità di risparmio desumibile dagli stessi redditi.

#### **4.2 Approccio stocastico**

I metodi stocastici si basano sulla modellazione probabilistica dei dati finalizzata ad individuare contribuenti che presentano caratteristiche sintomatiche di un possibile rischio fiscale, mediante un percorso logico inverso rispetto a quello tipico dei metodi deterministici.

Mentre, infatti, i modelli deterministici partono da un'ipotesi individuata a priori - che mette in relazione le informazioni fiscali, patrimoniali e finanziarie - e ne verificano la coerenza, i modelli stocastici ricercano, mediante l'impiego di tecniche statistiche, l'esistenza di profili segnaletici di comportamenti fiscalmente rischiosi, identificando gruppi di soggetti caratterizzati da relazioni anomale tra le variabili di interesse.

I due approcci possono essere adottati anche congiuntamente: gli esiti delle analisi statistiche possono, ad esempio, essere utilizzati per ridurre la platea di riferimento definita attraverso criteri deterministici, o per dare un ordine di priorità alle posizioni individuate.

In particolare, mediante un modello stocastico può stimarsi, per ciascun contribuente selezionato con un modello deterministico, la probabilità che l'esito di un'attività istruttoria nei suoi confronti dia luogo ad un'archiviazione (c.d. falso positivo); pertanto, tutti i soggetti per i quali tale probabilità

risultati superiori ad una determinata soglia, possono essere esclusi dalla platea iniziale di riferimento, in modo da incrementare l'efficacia della selezione.

Inoltre, le posizioni individuate possono essere ordinate in base alla stima della predetta probabilità di archiviazione, al fine di concentrare l'attività di controllo su quelle caratterizzate da una minore probabilità di essere archiviate e, quindi, da un maggior beneficio atteso.

In maggior dettaglio, i metodi di analisi probabilistica possono essere classificati in supervisionati e non supervisionati.

L'analisi supervisionata, detta anche predittiva, presuppone la disponibilità, su un campione di contribuenti (*training set*), di una variabile "risposta" (*output*) che rappresenta la grandezza che si intende predire per l'intera platea di soggetti. La variabile risposta può essere, ad esempio, la classificazione di un soggetto come "rischioso" o "non rischioso", oppure il maggior imponibile IRPEF accertato in esito ad un'attività di controllo.

Sulla base del *training set*, in una prima fase (detta di addestramento del modello), vengono analizzate le relazioni tra specifiche variabili fiscali, finanziarie, patrimoniali e dichiarative (variabili di *input*), disponibili su tutta la platea, e la variabile risposta; in una seconda fase, viene attribuita a ciascun soggetto della platea una previsione del valore della variabile risposta sulla base delle relazioni stimate (fase predittiva).

Tale processo non comporta il rischio di identificare gruppi di soggetti sulla base di criteri potenzialmente "discriminatori" in quanto né i dati di cui agli artt. 9 e 10 del Regolamento, né altre informazioni personali che potrebbero configurare un rischio (indiretto) di discriminazione (ad es. lo Stato di nascita) sono utilizzate come variabili esplicative dei modelli predittivi.

I metodi non supervisionati, a differenza di quelli supervisionati, non prevedono la presenza di una variabile di *output* nota a priori su un campione di soggetti, ma sono finalizzati ad individuare configurazioni dei dati associabili a comportamenti "rischiosi" da un punto di vista fiscale. Un

esempio di analisi non supervisionata può riguardare contribuenti per i quali, dall'analisi statistica delle distribuzioni delle variabili interessate, emerge un'anomala relazione tra consistenze patrimoniali risultanti dall'Archivio e i redditi dichiarati.

Si precisa che, anche in questa tipologia di metodi, non sono utilizzate variabili che potrebbero portare a classificazioni non rilevanti fiscalmente, come, ad esempio, la nazione di origine del contribuente.

Alla luce di quanto fin qui illustrato, l'applicazione delle metodologie in parola non determina in alcun modo la profilazione dell'intera popolazione dei contribuenti. L'utilizzo dei dati dell'Archivio, infatti, eventualmente interconnesso con altre banche dati nella disponibilità dell'Agenzia delle entrate, è volto unicamente alla selezione di un numero circoscritto di soggetti, caratterizzati da un apprezzabile livello di rischio fiscale.

Inoltre, le incongruenze oggettive e concrete che connotano la posizione fiscale di un contribuente, emerse nell'ambito delle analisi svolte, saranno oggetto di successivi approfondimenti a cura delle strutture locali preposte ai controlli, le quali garantiranno in ogni caso il pieno rispetto del principio del contraddittorio.

#### **4.3 Algoritmo di ottimizzazione del processo di selezione delle posizioni da sottoporre a controllo**

Di seguito viene illustrata la logica sottostante un algoritmo già implementato dall'Agenzia delle entrate per supportare gli Uffici locali preposti al controllo nell'attività di selezione delle posizioni da sottoporre ad attività istruttoria.

L'obiettivo dell'algoritmo è quello di consentire agli Uffici preposti al controllo di ordinare - secondo diversi criteri di priorità - le posizioni già individuate in esito ad un'analisi deterministica, così da ottimizzare i risultati e la calendarizzazione delle attività istruttorie.

L'algoritmo in questione è di tipo supervisionato e come *training set* è stata utilizzata la base dati riferita controlli effettuati dall'Agenzia delle entrate il cui esito è noto e cristallizzato.

Partendo, quindi, dai dati relativi ai contribuenti già oggetto di controllo, sono state individuate le informazioni fiscalmente rilevanti in grado di predire l'esito delle attività istruttorie, analizzate distintamente per tipologia della "fonte di innesco" (*i.e.*, motivi a base dell'inclusione del contribuente in una lista predisposta in base a specifici criteri di rischio, presenza di un processo verbale di constatazione, commissione di infedeltà dichiarative, coinvolgimento in schemi fraudolenti *et similia*).

Procedendo in tal senso, sono stati implementati due modelli:

- un primo modello finalizzato a predire, per ciascuna posizione in carico alla struttura territoriale, la probabilità che l'istruttoria dia esito negativo e che quindi la posizione venga archiviata;
- un secondo modello che, condizionatamente ad un esito positivo della fase istruttoria, e tenendo conto delle diverse tipologie di accertamento, produce una stima della probabilità di riscuotere gli importi accertati.

Da quanto esposto si evince che, mentre la prima fase della modellazione attiene alle decisioni prese dal personale preposto al controllo sulla base delle informazioni a sua disposizione, la seconda coinvolge invece anche le scelte effettuate dal contribuente in funzione della propria *propensione all'adempimento*. In particolare, il contribuente potrà reagire passivamente, oppure attivarsi in risposta alla notifica della pretesa da parte dell'Amministrazione fiscale. Nel primo caso, il percorso si concluderà tipicamente con l'esito "Ruolo per mancata impugnazione"; nel secondo caso, invece, l'accertamento potrà evolvere in contenzioso, essere annullato in autotutela, o si potrà definire mediante un istituto deflattivo (es. acquiescenza, accertamento con adesione, mediazione etc.) dell'originaria pretesa tributaria, con conseguente possibile decremento di quest'ultima sino a un importo "condiviso".

Si sottolinea nuovamente che lo schema descritto non determina in alcun modo l'emanazione di provvedimenti impositivi nei confronti di contribuenti selezionati mediante procedure totalmente automatizzate. Al contrario, secondo la logica descritta, il ruolo dei modelli di analisi è puramente ancillare, essendo limitato al supporto delle strutture di controllo nella gestione di posizioni che sono venute all'attenzione dell'Amministrazione finanziaria per motivi indipendenti dall'applicazione dei modelli stessi.

Infatti, come già rilevato, la variabile predittiva sarà associata esclusivamente ai soggetti per i quali è già stato deterministicamente individuato un rischio fiscale e non vincolerà in alcuno modo il personale addetto alle strutture di controllo.

## ALLEGATO TECNICO

### *ESEMPIO DI APPLICAZIONE DELLA LOGICA DEGLI ALGORITMI*

Nel presente allegato viene fornito un esempio di applicazione di un algoritmo di ottimizzazione di tipo stocastico, recentemente realizzato dall'Agenzia delle entrate. L'algoritmo in parola è stato sviluppato al fine di supportare gli Uffici preposti al controllo ad effettuare efficaci selezioni delle posizioni nei cui confronti avviare un'attività istruttoria.

In particolare, detto algoritmo consente di ordinare le diverse posizioni di soggetti già individuati a seguito di specifiche analisi del rischio, secondo un *rank* che può essere utilizzato per stabilire un criterio di priorità nella calendarizzazione delle attività. Ovviamente, tale elemento viene integrato dalle informazioni possedute dalle strutture di controllo e, di conseguenza, costituisce solo una delle componenti del quadro istruttorio complessivo.

Il metodo di analisi adottato è di tipo *supervisionato*: sulla base dell'insieme dei dati storici (*training set*) relativi ai controlli con esito definitivo (cioè non più modificabile), viene *addestrato* un modello predittivo che associa alle informazioni note *ex-ante* - relative a ogni posizione selezionata - la probabilità che a valle della fase istruttoria si registri un esito positivo del controllo.

La modellazione si articola in procedure separate, che corrispondono:

- a) alle diverse fasi temporali del controllo;
- b) alle diverse modalità istruttorie utilizzabili;
- c) a gruppi omogenei di contribuenti e di fonti di innesco (*i.e.*, motivi a base dell'inclusione del contribuente in una lista predisposta in base a specifici criteri di rischio, presenza di un processo verbale di constatazione, commissione di infedeltà dichiarative, coinvolgimento in schemi fraudolenti *et similia*).



Relativamente al punto c) si precisa, in particolare, che ad ogni gruppo omogeneo di contribuenti corrisponde uno specifico *set* di *parametri* del modello predittivo.

Di seguito viene fornito un esempio di applicazione dello schema descritto.

### ***ESEMPIO***

#### *Fase 1 – Individuazione di un rischio fiscale con una metodologia deterministica*

Tra le imprese individuali che operano nel settore del commercio al dettaglio, in un ambito merceologico che ipotizziamo essere tipicamente caratterizzato da elevati margini, vengono selezionate quelle che, relativamente all'anno  $t$ , manifestano le seguenti peculiarità:

- a) dall'analisi dei dati della fatturazione elettronica e dei corrispettivi emerge che la differenza tra i ricavi conseguiti nell'attività di impresa e i costi sostenuti è inferiore al 5% dei costi;
- b) il valore aggiunto calcolato sulla base delle dichiarazioni fiscali presentate risulta estremamente esiguo in relazione al numero di dipendenti dell'impresa (criterio applicato nel solo caso in cui l'impresa abbia personale dipendente);
- c) il valore della variabile *reddito d'impresa per addetto* è inferiore al decimo percentile della corrispondente distribuzione riferita alla totalità dei soggetti che operano nel medesimo settore economico.

Ai titolari delle ditte individuali identificati in base all'occorrenza congiunta delle condizioni a), b) e c) sono successivamente abbinati le informazioni desumibili dall'archivio dei rapporti finanziari. In particolare, per ciascuno di tali soggetti, si considera l'ammontare complessivo dei flussi in *avere* risultante, sempre nell'anno  $t$ , dai conti correnti e dagli altri rapporti finanziari.

Vengono quindi selezionati i soggetti per i quali detto ammontare risulta cumulativamente:

- superiore di almeno il 150% rispetto ai ricavi dichiarati;
- non inferiore a 300.000 €.

Infine, dalla platea così ottenuta sono espunti i contribuenti per i quali dai dati del Registro, ovvero dalle certificazioni dei sostituti di imposta, risultino atti che possano giustificare incrementi patrimoniali compatibili con le risultanze dell'archivio (vendita di immobili, donazioni e/o eredità ricevute, redditi sottoposti a tassazione sostitutiva, etc.).

Per quanto detto, il rischio fiscale oggetto del percorso di analisi testé delineato consiste essenzialmente nella possibile (o altamente probabile) sotto-dichiarazione dei ricavi conseguiti nell'attività d'impresa. Ai fini dell'individuazione dei soggetti rischiosi, sono infatti inclusi nell'analisi diversi indicatori che costituiscono potenziali segnali di occorrenza del fenomeno indagato (valori anomali di reddito e valore aggiunto, incoerenza tra ricavi dichiarati e flussi finanziari in *avere* registrati sui conti, etc.).

L'applicazione dei criteri idonei a intercettare i predetti segnali di rischio conduce alla predisposizione di una lista di nominativi da inviare alle Direzioni Provinciali (DP) competenti per le ordinarie attività di controllo.

Una volta trasmesse, le posizioni segnalate vengono gestite dalle predette DP che ne calendarizzano l'eventuale inserimento nei piani annuali di lavorazione.

Le informazioni relative alle ditte individuali che sono state selezionate in base al percorso illustrato, insieme a quelle riguardanti eventuali nuove posizioni emerse in esito ad altri percorsi di analisi, sono pre-elaborate da specifiche procedure che omogenizzano formati e codifiche in modo da renderle utilizzabili per le successive elaborazioni. Ai dati così standardizzati viene applicato l'algoritmo descritto nell'introduzione che, come anticipato, è stato in precedenza *addestrato* usando i dati storici delle posizioni già sottoposte a controllo.

In particolare, l'algoritmo opera attraverso i seguenti passaggi:

1. attribuzione di ciascuna nuova posizione all'appropriato gruppo di soggetti, in corrispondenza del quale è stato addestrato uno specifico modello (ad esempio, le posizioni emerse a valle del

percorso in argomento saranno incluse nella platea delle persone fisiche titolari di partita IVA);

2. applicazione di un modello predittivo che, utilizzando i parametri già stimati sulla base dei dati storici, attribuisce a ciascuna posizione la probabilità di archiviazione ovvero di attivazione di un accertamento;
3. applicazione di un modello predittivo che, condizionatamente all'attivazione di un accertamento, attribuisce, per ogni possibile modalità istruttoria, un livello di probabilità alle possibili reazioni del contribuente: versamento dell'importo corrispondente alla pretesa erariale; atteggiamento inerte con conseguenziale iscrizione a ruolo delle somme dovute etc.

La variabile “risposta” (variabile di *output*) di entrambi i modelli predittivi citati nei punti 1. e 2. è una variabile binaria (0/1) che codifica, rispettivamente, la realizzazione dell'evento “archiviazione” (o del suo complementare) e dell'evento “mancata impugnazione del provvedimento” (o del suo complementare). Le variabili esplicative dell'*output* (*predittori* o *covariate*) sono scelte tra quelle note *ex-ante* per tutte le posizioni che pervengono alla struttura, in modo da massimizzare il potere predittivo del modello, che viene misurato con opportuni indicatori di accuratezza.

Nel caso oggetto del presente esempio, in cui la platea è costituita da persone fisiche titolari di partita IVA, i predittori maggiormente significativi risultano essere i seguenti:

- fascia di età;
- settore di attività economica;
- volume d'affari;
- reddito complessivo;
- reddito d'impresa;
- reddito da partecipazioni;
- rapporto tra costi e ricavi;
- tipo di fonte d'innescò del controllo (verbale di constatazione, iniziativa dell'ufficio, etc);
- numero di accertamenti ricevuti in anni precedenti;

- numero di precedenti iscrizioni a ruolo;
- percezione di redditi da pensione;
- indicatori di ricchezza, quali numero di immobili e numero di automobili di proprietà;
- importo delle utenze per servizi.

Gli *output* dei modelli predittivi sopra menzionati sono combinati in un unico indicatore che, misurando il “valore” di ciascuna posizione in termini di rischiosità e proficuità, può essere utilizzato per assegnare un ordine di priorità alle posizioni stesse e ottimizzare il processo di calendarizzazione delle attività di controllo.

L'esempio mostra che il ricorso alla modellazione statistica dei dati non implica in alcun modo l'automatica emanazione di provvedimenti impositivi nei confronti dei contribuenti selezionati mediante procedure totalmente automatizzate. Al contrario, secondo la logica descritta, il ruolo dei modelli di analisi stocastica è puramente ancillare, essendo limitato al supporto delle strutture di controllo nella gestione di posizioni che sono venute all'attenzione dell'Amministrazione finanziaria per motivi indipendenti dall'applicazione dei modelli stessi. Tuttavia, i modelli così sviluppati consentono di fornire agli organi di controllo delle indicazioni che, in ultima analisi, potranno delineare con maggiore accuratezza i soggetti che presentano elevate probabilità di violazione di norme tributarie.