



OPENGOVWEEK  
19 - 23 maggio 2025

ITALIA  
●●●●● opengov

# Intelligenza Artificiale per l'integrità fiscale: modelli predittivi e open data nella selezione automatizzata dei rischi

# Contesto storico e rischi sistemici

La correlazione tra evasione fiscale e corruzione

Evasione e corruzione condividono radici sistemiche quali: l'opacità dei dati e la frammentazione dei controlli.

Le esperienze maturate durante il periodo di Tangentopoli evidenziano come irregolarità contabili all'interno delle imprese (pubbliche e private) abbiano favorito la creazione dei cd. «fondi neri», successivamente utilizzati per alimentare pratiche corruttive, in particolare nell'ambito delle procedure di appalto.

# PNRR e innovazione digitale nella fiscalità

Finanziamento europeo per modernizzare l'analisi dei rischi fiscali

L'Agenzia delle Entrate ha avviato una strategia di digitalizzazione dei flussi normativi, sostenuta da un finanziamento europeo di 900 MLN € nell'ambito del PNRR.

Il progetto, intitolato «A data driven approach to tax evasion risk analysis in Italy», promuove l'adozione di strumenti avanzati per l'elaborazione dei dati, integrati in un ecosistema 4.0 che modernizza i processi informativi dell'Amministrazione Finanziaria.

# IA e performance nelle indagini fiscali

Ottimizzazione dell'efficienza operativa

L'impiego di algoritmi predittivi e tecniche di machine learning consente di rafforzare l'efficienza operativa dei funzionari, ottimizzando i tempi di analisi, riducendo l'errore umano e potenziando l'accuratezza nella selezione dei contribuenti a rischio.

L'IA apprende dalle esperienze pregresse e migliora progressivamente le capacità decisionali, offrendo supporto in tempo reale e rendendo più mirate e tempestive le attività di accertamento.

# Network science e contrasto all'evasione

Tecniche avanzate per la ricostruzione di legami occulti tra soggetti

La network science applicata al fisco si fonda sulla network analysis, metodologia che consente di analizzare relazioni tra soggetti in forma di reti interconnesse, costituite da nodi e archi. Questa tecnica evidenzia collegamenti indiretti, che sfuggono alle tradizionali metodologie di indagine. La network analysis consente di intercettare triangolazioni fittizie e operazioni elusive, che costituiscono la base operativa per frodi sistemiche. Il valore aggiunto di questo approccio risiede nella capacità di processare dati in tempi contenuti, riducendo fabbisogno di risorse umane.

# Machine learning e performance operativa

Apprendimento dai dati per ottimizzare i controlli

Il machine learning migliora progressivamente la capacità di analisi, attivando impulsi operativi al superamento di soglie critiche determinate dall'AdE, sulla base della tipologia di indagine fiscale che si vuole implementare.

I modelli si evolvono imparando dalle attività pregresse, grazie anche a meccanismi di deep learning, supportando i funzionari con maggiore efficienza, riducendo drasticamente la percentuale di errore.

# IA simbolica e non simbolica

Algoritmi deterministici vs probabilistici

L'IA si articola, in due famiglie di algoritmi: simbolici e non simbolici, differenziati per logica operativa e grado di determinismo. Gli algoritmi simbolici, si fondano su una struttura rigida, costruita su sequenze causali “if-then-else”. Questo approccio è il punto di partenza dell'IA e si presta a contesti normativi dove la trasparenza è essenziale. Gli algoritmi non simbolici adottano una logica probabilistica, modellando la realtà attraverso il calcolo delle probabilità, stimando quale output sia più plausibile sulla base dell'esperienza pregressa (reti neurali, il machine learning, il deep learning).

# Apprendimento supervisionato

Modelli basati su dati con output noto

L'apprendimento supervisionato è una metodologia dell'IA, in cui il sistema viene istruito attraverso training set. Il modello apprende le relazioni esistenti tra variabili e generalizza regole decisionali applicabili a nuovi casi simili. La dipendenza da un dataset iniziale rappresenta una criticità: la fase di progettazione e codifica richiede tempi lunghi, inoltre un dataset distorto porta all'insorgenza di bias algoritmici. Per mitigare questi rischi, sono state sviluppate varianti random forests, che combinano l'output di più alberi decisionali per aumentare l'affidabilità complessiva.

# Data visualization per l'analisi fiscale.

Visualizzazioni interattive per il supporto decisionale

La data visualization è uno strumento che interpreta volumi di dati complessi, attraverso rappresentazioni grafiche. L'impiego di dashboard dinamiche, consente ai funzionari dell'AdE di orientarsi all'interno di sistemi informativi articolati, evidenziando, pattern irregolari e comportamenti fiscali atipici. L'interazione in tempo reale consente di generare analisi differenziate, aumentando la tempestività dell'intervento. La rappresentazione grafica, integrata con sistemi di IA, consente un'analisi predittiva immediata, agevolando la comunicazione tra team ispettivi.

# Reti neurali e deep learning

Apprendimento adattivo e complessità dei modelli algoritmici

Le reti neurali sono costituite da modelli computazionali, che apprendono dai dati attraverso un processo iterativo basato sull'esperienza. I sistemi sono addestrati fornendo coppie di input e output noti: analizzando gli errori commessi in fase di elaborazione, si correggono autonomamente grazie alla backpropagation, che modifica i pesi assegnati ai nodi della rete. Le reti neurali composte da strati intermedi ("hidden layers"), combinate con il deep learning, consentono di affrontare problemi estremamente complessi, come la modellazione predittiva avanzata.

# Apprendimento non supervisionato

Clusterizzazione dei dati per la scoperta di pattern latenti

L'apprendimento non supervisionato, tecnica di machine learning in cui l'IA analizza dati grezzi, è in grado di individuare correlazioni nascoste, suddividendo gli elementi in cluster. Utile per la scoperta di profili anomali, l'approccio non supervisionato è più economico e flessibile, in quanto non richiede la costruzione preventiva di dataset. Presenta minor rischio di bias, poichè l'algoritmo costruisce le categorie a partire dai dati osservati. L'affidabilità dipende dalla corretta interpretazione dei cluster generati, validati da analisti per evitare sovrainterpretazioni.

# Ve.Ra. e l'individuazione del rischio fiscale

Analisi automatizzata sui rapporti finanziari

Ve.Ra. (Verifica dei rapporti finanziari) è uno strumento di data analysis sviluppato per supportare l'individuazione dei rischi di evasione fiscale, basandosi sui dati dell'Archivio dei Rapporti Finanziari, in attuazione della legge di bilancio 2020 (art. 1, commi 681-686, l. 160/2019).

Il cuore del processo risiede nella capacità degli algoritmi di selezionare, all'interno di una vasta popolazione di contribuenti, un campione mirato di soggetti a rischio, guidati dalle caratteristiche peculiari del rischio da intercettare.

# Processo di analisi del rischio fiscale

Le cinque fasi per una selezione mirata dei contribuenti a rischio

L'AdE ha implementato un modello di analisi del rischio fiscale articolato in 5 fasi, basato su una combinazione di dati amministrativi, finanziari e algoritmi predittivi. Il processo si apre con una selezione iniziale su dati A.T., per poi integrarsi con le informazioni dell'Archivio Ve.Ra. Segue una fase di scrematura delle posizioni giustificabili, la trasmissione alle D.P. per l'affinamento e l'eventuale avvio di azioni ispettive o di compliance. L'intero sistema è costruito per garantire la centralità dell'intervento umano, evitando automatismi sanzionatori.

# Un sistema senza automatismi sanzionatorie

L'IA come supporto, non come giudice

I modelli predittivi non determinano provvedimenti automatici.

Il loro ruolo è ancillare, a supporto delle strutture di controllo umane.

Durante il processo di analisi, i dati personali dei contribuenti sono “pseudonimizzati”, 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.

# Recupero fiscale 2024: risultati e composizione

Incremento gettito fiscale del 6,5% rispetto al 2023

Nel 2024, seppur in assenza di statiche ufficiali, l'Agenzia delle Entrate ha recuperato complessivamente 26,3 miliardi di euro, con un aumento di 1,6 miliardi rispetto all'anno precedente (+6,5%).

Di questi, 22,8 miliardi derivano da attività di controllo ordinario.

# Open data, etica e innovazione

Verso una PA trasparente e inclusiva

Gli open data, accessibili in formato aperto e interoperabile tramite piattaforme come Dati.gov.it , rappresentano una leva strategica. Tuttavia, l'adozione dell'IA deve essere guidata da un solido impianto etico, giuridico e organizzativo: formazione dei funzionari, audit algoritmici e organismi indipendenti di supervisione sono strumenti imprescindibili per un uso responsabile delle tecnologie. L'esperienza dell'AdE costituisce un laboratorio avanzato di innovazione pubblica, da estendere ad altri ambiti dell'amministrazione, in una prospettiva di governance integrata.



**OPENGOVWEEK**  
**19 - 23 maggio 2025**

@opengovitaly | #OpenGovWeek | open.gov.it

**ITALIA**  
●●●●● OPENGOV