



PROFILAZIONE FISCALE E PRIVACY: UN COMPROMESSO POSSIBILE (?)

A.SANTORO, UNIVERSITÀ DEGLI STUDI DI MILANO BICOCCA

INPS, 18 NOVEMBRE 2019

OUTLINE

- ❑ PROFILAZIONE E PREDIZIONE
- ❑ UN ESEMPIO DI PROFILAZIONE FISCALE: BATTISTON, GAMBA AND SANTORO «OPTIMIZING TAX ADMINISTRATION POLICIES WITH MACHINE LEARNING» (2019)
- ❑ UN ALTRO ESEMPIO DI PROFILAZIONE FISCALE: L'USO DELL'ANAGRAFE DEI RAPPORTI
- ❑ LEGGE DI BILANCIO 2020 E PRIVACY

PROFILAZIONE FISCALE: UNA DEFINIZIONE

- *Operazione che associa ad insiemi di caratteristiche individuali e di variabili economiche e sociali **una certa probabilità di aver adottato in passato** o di **poter adottare in futuro** comportamenti fiscali rischiosi (evasione, elusione, mancato pagamento di imposte dichiarate, comunicazione di dati artefatti...)*

PROFILAZIONE ↔ PREDIZIONE

PREDIZIONE E POLICY FISCALI

- I **policy prediction problems** (Kleinberg et al., 2015) sono sempre più frequentemente studiati in modo autonomo/complementare rispetto ai tradizionali policy evaluation problems.
- *In Belgium, tax authorities have developed different predictive models that resulted in reducing the tax debt. These models are used to predict if a person or a company will pay his debt within 14 days after receiving a phone call and, in the case he did not pay, to predict if he will pay after sending a bailiff.*
- *In Canada, a comprehensive dataset is formed linking taxpayers' filing and assessment information, risk profiles, historical audits, collections and appeal information. Then data mining and machine learning algorithms are used to predict income and sales tax that small and medium enterprises should pay and finally attributing a risk score to each taxpayer*

UN ESEMPIO DI PROFILAZIONE FISCALE

(BATTISTON, GAMBA AND SANTORO, 2019)

Banca dati
Agenzia delle
Entrate

- Dati delle **dichiarazioni**: informazioni personali
- Dati sugli **studi di settore**: congruità, coerenza, soglia presunta

Predizione e
profilazione

- Profili individuali a cui è statisticamente associata una certa **probabilità** di dichiarare un ricavo esattamente pari alla soglia presunta.

Policy

- **Selezione dei contribuenti** a cui far sapere che il ricavo da dichiarare è quello vero, non quello presunto.

APPROCCIO METODOLOGICO/1

(BATTISTON, GAMBA AND SANTORO, 2019)

- Osserviamo al tempo t e al tempo $t+1$ un insieme di caratteristiche individuali e di variabili economiche a livello individuale.
- Utilizziamo diverse metodologie per **stimare la probabilità che al tempo $t+1$ venga adottato il comportamento a rischio** in funzione delle caratteristiche individuali, delle variabili economiche e delle possibili interazioni osservate al tempo t .
- Ad ogni profilo osservato al tempo t (insieme di caratteristiche individuali, valori di variabili economiche e delle possibili interazioni) è associabile una determinata probabilità di adozione del comportamento a rischio al tempo $t+1$.

APPROCCIO METODOLOGICO/2

(BATTISTON, GAMBA AND SANTORO, 2019)

- La policy viene applicata a tutti quei profili per i quali la probabilità di un comportamento a rischio è superiore ad un determinato valore o threshold (τ).
- Ad ogni policy sono associati due errori di predizione:
 - ✓ **falsi positivi:** contribuenti a cui VIENE applicata la policy al tempo $t+1$ anche se il comportamento osservato al tempo $t+1$ NON è a rischio;
 - ✓ **falsi negativi:** contribuenti a cui NON VIENE applicata la policy al tempo $t+1$ anche se il comportamento osservato al tempo $t+1$ E' a rischio.

APPROCCIO METODOLOGICO/3

(BATTISTON, GAMBA AND SANTORO, 2019)

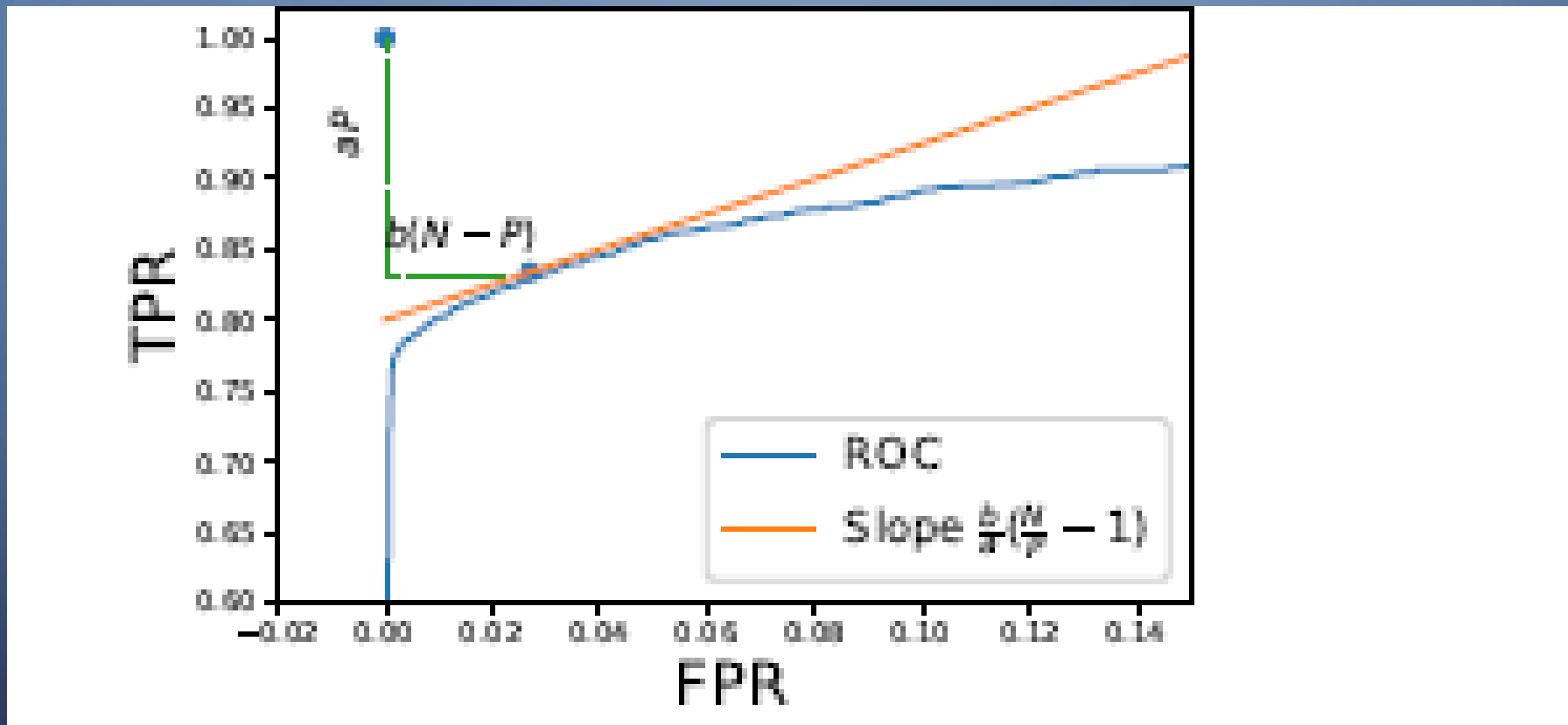
- Va quindi trovata una regola che consenta di **minimizzare la combinazione degli errori** di predizione tenendo conto del loro costo sociale.

$$\text{MIN } L = a\text{FNR}(\tau)P + b\text{FPR}(\tau)[N - P]$$

- dove
 - ✓ a = costo medio di ogni falso negativo (mancato recupero di risorse pubbliche al netto delle risorse pubbliche e private spese)
 - ✓ b = costo medio di ogni falso positivo (spreco di risorse pubbliche e private)
 - ✓ N = numero contribuenti, P = contribuenti che al tempo $t+1$ hanno comportamenti a rischio

POLICY OTTIMALE

(BATTISTON, GAMBA AND SANTORO, 2019)



UN ALTRO ESEMPIO DI PROFILAZIONE FISCALE

Matching delle banche dati su base individuale

- Dati delle **dichiarazioni**: informazioni personali
- Dati dell'**anagrafe dei rapporti**: informazioni finanziarie
- Dati sui **controlli**: esiti dei controlli (positivi/negativi)

Profilazione

- Profili individuali a cui è statisticamente associata una certa **probabilità** di avere evaso le imposte, ovvero di essere controllato con esiti positivi

Policy

- **Selezione dei controlli** sulla base della profilazione, tenendo conto della natura probabilistica del legame emerso

USO DELL'ANAGRAFE DEI RAPPORTI FINANZIARI: IL PASSATO RECENTE

- **Profilazione** utilizzando i dati dell'anagrafe dei rapporti finanziari teoricamente possibile dal **2011** [«...»].
- **Matching** dei dati individuali effettuato a livello di **singoli casi ed ex post** (indagini finanziarie) **non massivo ed ex ante**.
- **Relazione della Corte dei Conti 2017** conclude che Anagrafe dei rapporti è sostanzialmente non utilizzata.

PROFILAZIONE E PRIVACY: ARTICOLO 22 GDPR

- 1. L'interessato ha il diritto di non essere sottoposto a una decisione basata unicamente sul trattamento automatizzato, compresa la profilazione, che produca effetti giuridici che lo riguardano o che incida in modo analogo significativamente sulla sua persona.*
- 2. Il paragrafo 1 non si applica nel caso in cui la decisione: (...)*
 - b) sia autorizzata dal diritto dello Stato membro cui è soggetto il titolare del trattamento, che precisa altresì misure adeguate a tutela dei diritti, delle libertà e dei legittimi interessi dell'interessato.*

USO DELL'ANAGRAFE DEI RAPPORTI: L'INTERVENTO NORMATIVO (ARTICOLO 86,1 DDL BILANCIO 2020)

- Con riferimento alle informazioni contenute nell'**archivio dei rapporti finanziari** (...) l'Agenzia delle entrate, previa **pseudonimizzazione** dei dati personali, si avvale delle tecnologie, delle elaborazioni e delle interconnessioni con le altre banche dati di cui dispone, **allo scopo di individuare criteri di rischio utili per far emergere posizioni da sottoporre a controllo e incentivare l'adempimento spontaneo**, nel rispetto di quanto previsto dall'articolo 23 del Regolamento GDPR

USO DELL'ANAGRAFE DEI RAPPORTI: LA MEMORIA DEL GARANTE

- A proposito delle misure di sicurezza, il Garante suggerisce («sarebbe utile») di inserire un richiamo all'articolo 22, 2,b) e specifica che queste misure potrebbero consistere in:
 - controlli sulla **qualità** dei dati;
 - controlli sulle **elaborazioni logiche**;
 - cautele (...) con particolare riferimento ad **erronee rappresentazioni della capacità contributiva**.

QUALI «ELABORAZIONI LOGICHE»?

- La profilazione come prospettata in precedenza si basa su un ragionamento **induttivo**: non vi sono criteri stabiliti a priori, perché i criteri si evincono dai dati. Il legame tra **profilo e probabilità potrebbe non avere alcuna «logicità»** [è già difficile interpretare quello che c'è nelle black box, figuriamoci trovare un senso logico!].
- «Logiche» nel senso di **appropriate** devono essere le metodologie usate per stimare la probabilità e «logiche» nel senso di **robuste** dal punto di vista statistico devono essere le correlazioni trovate.

COSA SI INTENDE PER «ERRONEE RAPPRESENTAZIONI DELLA CAPACITA' CONTRIBUTIVA»?

- La profilazione punta a misurare una possibile sottorappresentazione della capacità contributiva. **Una certa frequenza di errori di predizione è quindi scontata.** Si può agire sulla minimizzazione della L, eventualmente sovrappesando i falsi positivi.
- Trattandosi di un metodo probabilistico, può essere utilizzato solo come **guida ai controlli**, e quindi l'esatta rappresentazione della capacità contributiva può essere ottenuta solo in fase applicativa.

LA QUESTIONE DELLE COMPETENZE

- **Se** si accetta di impostare la profilazione come esito di un'attività di analisi massiva dei dati, diventa fondamentale un controllo sulla qualità dei dati e delle **metodologie** applicate, nonché sulla **robustezza** dei risultati e sulla **misurazione** dell'errore di predizione.
- Ne segue che **Agenzia e Garante della privacy** vanno dotati delle **competenze statistiche ed econometriche necessarie**.