

AUDIZIONE DEL PRESIDENTE DEL GARANTE PER LA PROTEZIONE DEI DATI PERSONALI  
ANTONELLO SORO

L'UTILIZZO DELLE METODOLOGIE DI *DATA MINING* PER ESEGUIRE VISITE MEDICHE DI CONTROLLO AI  
LAVORATORI DEL SETTORE PUBBLICO  
COMMISSIONE LAVORO, PREVIDENZA SOCIALE  
SENATO DELLA REPUBBLICA 18 SETTEMBRE 2018

L'introduzione di un sistema di programmazione mirata delle visite fiscali da parte dell'Inps, volto a individuare preventivamente possibili assenze ingiustificate dal lavoro per malattia, costituisce indubbiamente un obiettivo di interesse generale.

Indirizzando i controlli verso i casi maggiormente suscettibili di rivelare abusi, infatti, tale procedura promuoverebbe l'efficienza del sistema previdenziale e un risparmio anche rilevante di risorse.

Il perseguimento di tale obiettivo mediante un modello statistico predittivo, basato sull'analisi della probabilità di comportamenti fraudolenti da parte dei lavoratori o di eventi di malattia risolvibili prima del previsto, offre la possibilità di concentrare i controlli sui casi considerati maggiormente a rischio di abusi.

Il sistema adottato dall'Inps, in particolare, è basato su processi decisionali automatizzati consistenti in una profilazione dei lavoratori, che attribuisce alle certificazioni di malattia un punteggio di maggiore o minore affidabilità su base statistica del giudizio prognostico di idoneità alla ripresa del lavoro o di assenza ingiustificata.

Il ricorso a tali tecniche, anche per scopi del tutto legittimi quali – nella fattispecie – la prevenzione di comportamenti fraudolenti, è ammesso dalla disciplina sulla protezione dei dati personali in circostanze specifiche, purché risponda a criteri di correttezza e trasparenza e osservi sufficienti garanzie, soprattutto quando riguarda dati sulla salute.

Tali garanzie dovrebbero, innanzitutto, consentire alle persone interessate di essere informate dell'esistenza di un trattamento automatizzato suscettibile di riguardarle, in modo da poter verificare e, ove necessario, contestare i risultati che ne derivino.

In secondo luogo, dovrebbero essere adottate misure volte a ridurre i margini di errore insiti in queste procedure di elaborazione statistica – caratterizzate dall'applicazione automatica di regole prestabilite di inferenza – e, quindi, i rischi di giudizi prognostici errati, nonché di abusi, di discriminazioni ingiustificate e, soprattutto, di violazione dei diritti degli interessati.

Più in generale i processi decisionali automatizzati, ivi inclusa la profilazione, sono utilizzati in un numero crescente di settori, tanto del privato (sulla base, tra gli altri presupposti, del consenso esplicito dell'interessato) quanto del pubblico. Banche e finanza, assistenza sanitaria, fiscalità, assicurazioni, marketing e pubblicità sono soltanto alcuni dei campi nei quali la profilazione viene effettuata con maggiore regolarità a sostegno del processo decisionale.

L'Ufficio del Garante è intervenuto sulle modalità di programmazione delle visite di controllo adottate dall'Inps nel febbraio di quest'anno, dopo aver appreso da alcune notizie di stampa dell'elaborazione di un sistema di *data mining*, denominato SAVIO, operante non solo all'insaputa dei lavoratori interessati, ma anche in assenza di precauzioni e garanzie specifiche

volte ad evitare, ad esempio, che inesattezze nei dati raccolti o incongruenze nella logica degli algoritmi utilizzati, inducano decisioni erranee con impatti negativi sui singoli.

Nel corso dell'istruttoria abbiamo appreso che l'Inps si sarebbe avvalso da oltre cinque anni di un sistema di programmazione mirata delle visite fiscali basato su un trattamento automatizzato di dati personali che, a seguito della creazione del Polo unico della medicina fiscale, sarebbe stato esteso anche ai pubblici dipendenti.

L'Istituto aveva infatti ommesso di notificare all'Autorità – come all'epoca previsto – la circostanza dell'adozione di tale meccanismo predittivo di *data mining*, basato sulla profilazione. Come precisato dall'Istituto, questo non aveva ritenuto che il sistema SAVIO consistesse in un trattamento di profilazione volto ad individuare i lavoratori per i quali la condizione o la prognosi di malattia non corrispondesse del tutto alla condizione reale.

L'Autorità quindi – oltre ad offrire la propria piena disponibilità a collaborare per realizzare un sistema di programmazione delle visite fiscali efficiente tanto quanto rispettoso del diritto alla protezione dei dati personali – ha rilevato l'omissione in cui è incorso l'Istituto e, soprattutto, l'assenza delle necessarie misure di garanzia.

Ne è seguita la decisione dell'Istituto di sospendere l'attività di profilazione a fini di programmazione delle visite fiscali.

In sede di audizione, il Presidente dell'Inps ha sottolineato come tra le variabili considerate nel modello SAVIO non vi sia la diagnosi, ossia la patologia da cui è affetto il lavoratore, riportata nel certificato medico.

Il software predittivo elaborato prenderebbe, infatti, in considerazione soltanto la frequenza e la durata dei singoli episodi di malattia del lavoratore, insieme ad altre variabili quali il numero delle precedenti idoneità alle visite mediche di controllo, la qualifica del lavoratore, il tipo di rapporto di lavoro, la retribuzione, il settore e la dimensione aziendale.

Tuttavia, il dato stesso dell'assenza dal lavoro per malattia costituisce, per ciò solo, – come da consolidato orientamento del Garante, confermato anche dalla giurisprudenza di legittimità – dato sulla salute e, come tale, meritevole della tutela rafforzata che l'ordinamento europeo e le norme interne accordano a tale categoria di dati sensibili.

Tanto più se accanto a tali informazioni, come quelle prese in considerazione dal software predittivo elaborato dall'Inps, vi siano i dati riferiti alla frequenza e alla durata dei singoli episodi di malattia del lavoratore e alla storia pregressa delle stesse, risultanti in particolare dai certificati medici prodotti negli ultimi due anni, nonché dagli esiti delle precedenti visite di controllo.

Il modello statistico di analisi in questione, contrariamente a quanto ritenuto, realizza, quindi, una vera e propria profilazione dei lavoratori interessati.

In base alle indicazioni contenute nella raccomandazione del Consiglio d'Europa in materia di profilazione e alle previsioni del nuovo regolamento europeo sulla protezione dei dati, questa è una tecnica che, correlando un certo numero di informazioni individuali, consente di inserire il singolo in una categoria predeterminata, il "profilo" appunto, costruito sulla base di alcune variabili o caratteristiche comportamentali che accomunano i soggetti ricompresi nella categoria di riferimento.

La ricomprensione dell'individuo all'interno di un determinato profilo ha, tuttavia, un effetto non soltanto descrittivo ma anche predittivo ed ascrittivo, in quanto ne fa desumere, sulla base di un'inferenza statistica (peraltro suscettibile di errore), la propensione alla realizzazione di specifici comportamenti futuri.

Nella fattispecie, la profilazione dei lavoratori determina l'attribuzione a ciascuno di un determinato grado di propensione all'assenza per malattia ingiustificata.

È bene chiarire che le attività di analisi statistica realizzate in tal modo sui certificati medici, e su altre variabili riferite al lavoratore, non consistono in una mera selezione degli individui sulla base delle loro caratteristiche reali: attività questa, che da un punto di vista tecnico, non richiede il *data mining*.

Tali analisi determinano, invece, un'effettiva profilazione, nella misura in cui, sulla base di inferenze statistiche, realizzano una prognosi comportamentale, che in quanto fondata su un calcolo probabilistico è sempre soggetta a un margine di errore e per questo necessita di garanzie adeguate per evitare false attribuzioni e valutazioni erranee dei comportamenti individuali.

Per tali ragioni, il nuovo regolamento europeo tratta in maniera specifica la profilazione e il processo decisionale automatizzato relativo alle persone fisiche (e non delle persone giuridiche che, come noto, non beneficiano della tutela accordata dalla normativa di protezione dati, se non nel settore delle comunicazioni elettroniche).

Tale disciplina, ha introdotto nuove disposizioni per far fronte ai rischi derivanti dall'utilizzo di tali tecniche, in relazione alla tutela della vita privata.

Nello specifico, il regolamento vieta di utilizzare dati sensibili, ivi compresi quindi quelli sulla salute, in processi decisionali completamente automatizzati che incidano in modo significativo sugli interessati a meno che tali attività non siano necessarie per motivi di interesse pubblico rilevante, sulla base del diritto nazionale vigente (che deve comunque prevedere misure adeguate e specifiche a tutela dei diritti delle libertà e dei legittimi interessi delle persone coinvolte), nonché essere proporzionato alla finalità perseguita, rispettando altresì l'essenza del diritto alla protezione dei dati personali.

In questo quadro, pertanto, pur non essendo in discussione la rilevanza e la legittimità degli scopi perseguiti mediante tale sistema dall'Inps, è imprescindibile che vengano adottate misure di salvaguardia adeguate a tutela degli interessati in conformità al nuovo quadro giuridico europeo.

Al riguardo, non è peraltro sufficiente un intervento normativo che autorizzi semplicemente il ricorso a tali tecniche da parte dell'Istituto, individuando gli interessi pubblici rilevanti per i quali queste possono essere impiegate.

È, infatti, necessario che la normativa interna introduca regole puntuali volte ad accordare le garanzie necessarie per il rispetto dei diritti dei lavoratori interessati.

La proposta normativa elaborata dall'Inps, così come formulata, non pare idonea a conseguire l'obiettivo auspicato.

Si tratta, infatti, di una norma quadro di portata generale che non soddisfa i requisiti richiesti dal diritto europeo, in quanto si limita ad enunciare in modo generico gli interessi pubblici per i quali l'Istituto sarebbe autorizzato ad impiegare processi decisionali automatizzati con dati personali anche sensibili, senza circoscrivere in modo chiaro e inequivocabile le rilevanti finalità che possano ritenere giustificato, rispetto agli specifici obiettivi perseguiti, il ricorso a tali tecniche, né approntare le necessarie garanzie per i lavoratori interessati.

In particolare, dovrebbero essere previste le modalità con cui informare gli interessati dell'esistenza di una profilazione, dei criteri utilizzati, nonché delle conseguenze derivanti dall'attribuzione di un determinato profilo alla propria situazione individuale.

Dovrebbero essere anche adottate misure tecniche e organizzative appropriate che consentano di rettificare eventuali fattori tali da incidere sull'esattezza dei dati e garantire il loro aggiornamento costante, evitando così che i profili vengano costruiti o attribuiti ai singoli sulla base di dati imprecisi.

I dati dovrebbero essere conservati per un tempo proporzionato.

Inoltre, la pertinenza e la qualità delle informazioni e delle inferenze statistiche utilizzate andrebbero rivalutate periodicamente e in un tempo ragionevole.

Gli algoritmi utilizzati andrebbero configurati in modo da minimizzare il rischio di errori o distorsioni .

Particolare attenzione andrebbe prestata alla minimizzazione dei dati, specie nel caso di correlazioni statistiche tra informazioni relative a sfere diverse della di vita privata degli interessati. Da questo punto di vista sarebbe, ad esempio, opportuno giustificare, sulla base di elementi obiettivi e ragionevoli, l'inserimento nel modello statistico dei dati sulla retribuzione, sulla qualifica professionale, sul settore e sulla dimensione aziendale o anche l'identificativo del medico che ha redatto il certificato.

Gli interessati dovrebbero avere a disposizione un modo semplice per esercitare i loro diritti e, in particolare, quello di accedere alle informazioni utilizzate e di correggere eventuali informazioni errate.

Infine, il titolare del trattamento dovrebbe effettuare una valutazione d'impatto, in modo da valutare adeguatamente, in concreto, i rischi connessi al processo decisionale automatizzato e le misure necessarie in funzione preventiva.

Tali previsioni (che potrebbero essere inserite anche in un atto normativo secondario al quale la fonte primaria faccia rinvio) consentirebbero di realizzare un sistema idoneo a coniugare, nella maniera migliore, l'efficienza degli accertamenti sulle assenze dal lavoro e il diritto alla protezione dei dati personali, che rappresenta in misura sempre crescente un essenziale presidio di libertà e dignità nella società digitale.