

Rapporto 1/2021 – L'impiego dell'intelligenza artificiale nell'attività di CONSOB, AGCOM e ARERA

Edoardo Chiti

Professore ordinario di Diritto amministrativo, Università della Tuscia e Sant'Anna di Pisa

Barbara Marchetti

Professoressa ordinaria di Diritto amministrativo, Università degli Studi di Trento

Nicoletta Rangone

Professoressa ordinaria di Diritto amministrativo, Università Lumsa

L'uso dell'intelligenza artificiale nelle autorità amministrative indipendenti: applicazioni concrete e prospettive

Il seminario che ha coinvolto Consob, Agcom e ARERA si è incentrato su poche questioni chiave: quali applicazioni o sperimentazioni dell'intelligenza artificiale sono portate avanti dalle autorità indipendenti? Che tipo di tecnologia è interessata? Quali dati la alimentano? Chi governa e supervisiona il processo?

Ne è emerso che Consob ha optato per un modello organizzativo che consente un approccio sistemico all'innovazione, che guarda alle evoluzioni dei mercati, sia per dare una risposta di vigilanza adeguata (incubatore di *digital innovation*), sia per supportare l'innovazione nei mercati (*sandboxes* e *fintech*). Le sperimentazioni in atto sul versante vigilanza attengono, quanto ai prodotti finanziari, ai controlli sui KID e PRIIPs informati al rischio e basati sul *machine learning* e, in ordine agli abusi di mercato, all'*intelligent crawling* per individuare le piattaforme *on line* attraverso le quali vengono svolte abusivamente attività riservate.

Di particolare interesse è anche l'attività svolta dall'Autorità per le garanzie nelle comunicazioni per rilevare la disinformazione *on-line* attraverso un monitoraggio sistematico di articoli di quotidiani e siti web, trascrizioni di trasmissione televisive e radio, tweet/post pubblicati su piattaforme, i cui risultati sono pubblicati dall'Osservazione sulla disinformazione *on-line* appositamente istituito.

Variegata infine l'esperienza di ARERA che, pur non facendo uso di sistemi di intelligenza artificiale in senso stretto, ha evidenziato tre livelli progressivi di utilizzo di intelligenza nell'analisi dei dati: a un primo livello, la messa a disposizione della struttura di strumenti di "intelligenza assistita" (per l'analisi automatica o semi automatica di dati) ha consentito di ridurre notevolmente il tempo necessario alla "preparazione" dei dati necessari all'attività di regolazione e monitoraggio. A un secondo livello, sono stati messi a punto meccanismi di "intelligenza condivisa" con l'Agenzia europea per la collaborazione dei regolatori dell'energia (ACER), finalizzati al monitoraggio dei mercati energetici all'ingrosso su scala europea, in particolare per selezionare casi "sospetti" dietro i quali potrebbero nascondersi abusi di mercato, da sottoporre ad approfondimento "umano". Infine, a un terzo livello che è stato definito "intelligenza della filiera", si è evidenziato come interventi di regolazione che aumentano la disponibilità di dati elementari, come il passaggio progressivo ai contatori elettrici di seconda generazione (*smart meter 2g*) con maggiore granularità temporale e minore latenza dei dati di consumo raccolti, ha consentito non solo agli operatori di rete (che sono titolari dell'attività di *metering*), ma anche ai soggetti di mercato di avere a disposizione una importante mole di dati sui consumi che possono essere elaborati con sistemi di intelligenza artificiale, in ultima analisi a potenziale vantaggio dei consumatori.

La discussione delle esperienze descritte ha portato all'emergere delle seguenti evidenze e criticità.

a) L'intelligenza artificiale tende ad essere utilizzata soprattutto per estrarre informazioni dai dati, al fine di migliorare le funzioni tradizionali delle autorità amministrative indipendenti. Le sperimentazioni riportate sembrano comunque fare riferimento ad utilizzi limitati alla raccolta di informazioni in fase preistruttoria (Stefano Civitarese Matteucci). Quand'anche l'intelligenza artificiale supporti l'attività di vigilanza, la decisione di aprire o meno un'istruttoria (che si tratti di un procedimento cautelare o sanzionatorio) resta una decisione umana (Giulia Avanzini; Giuseppe Frega, Consob).

b) Le autorità (come tutte le amministrazioni italiane) sperimentano e utilizzano strumenti di intelligenza artificiale al di fuori di un quadro normativo che definisca l'uso all'interno del procedimento e il ruolo che deve rivestire il decisore umano, come invece è stato in Francia con riferimento all'utilizzo di algoritmi da parte dell'amministrazione fiscale (Giulia Avanzini).

c) Nella maggior parte dei casi, le sperimentazioni e gli usi sono sviluppati con competenze interne alle autorità indipendenti, così quanto a Banca d'Italia e Consob, mentre ARERA utilizza anche prodotti o sistemi che non sono auto-prodotti (Giulia Avanzini). A quest'ultimo riguardo è interessante notare come l'interazione con il fornitore di nuove tecnologie non venga percepita come molto diversa da quella con qualunque altro fornitore (Luca Lazza, ARERA) e che le scelte, tanto della tecnologia quanto dello stesso fornitore, siano decisioni proprie degli uffici in quanto relative alla strumentazione utile all'istruttoria, laddove il collegio si limita a verificare che i risultati attesi siano stati raggiunti (Luca Lo Schiavo, ARERA).

d) Frequenti sono le collaborazioni con università e centri di ricerca che assicurano il necessario approccio interdisciplinare, che difficilmente si può raggiungere attraverso le sole competenze interne alle amministrazioni (Paola Deriu, Consob).

e) Tutti i progetti di uso dell'intelligenza artificiale nascono dalle esigenze degli uffici e si sviluppano grazie alla collaborazione tra chi ha una concreta esperienza di vigilanza e i *data scientists*, sia che lo strumento tecnologico venga sviluppato all'interno dell'amministrazione, sia che venga acquisito all'esterno. L'intelligenza artificiale è infatti considerata uno strumento abilitante, le cui funzionalità si sviluppano solo attraverso una "personalizzazione" effettuata in virtù dell'adattamento alle esigenze dell'amministrazione e la lavorazione dei dati forniti dall'amministrazione stessa (Luca Lazza, ARERA). Dunque non esiste un prodotto *off-the-shelf* che consenta di essere utilizzato, ad esempio, per l'individuazione tanto degli abusi di mercato di competenza di Consob, che di quelli sui cui interviene ARERA. Questo è fondamentale anche per consentire l'approccio modulare (adottato da Consob), volto a perseguire la spiegabilità dell'algoritmo da parte dell'amministrazione (Paola Deriu, Consob).

1. L'approccio SMART di Consob alla digitalizzazione finanziaria

Tiziana Togna

Vicedirettore generale CONSOB

La Consob ha optato non già per la creazione di singole unità dedicate alla digitalizzazione finanziaria, quanto per lo sviluppo di più ambienti in cui questa è studiata e sperimentata.

Il fulcro dell'azione di digitalizzazione è costituito dallo *Steering Committee Fintech*, costituito dalla Commissione e affidato al coordinamento di uno dei membri del collegio, il dottor Paolo Ciocca. Il *Committee* è nato e continua ad esistere come strumento di coordinamento di un'attività di studio e per la formulazione di proposte progettuali in materia di *Fintech*. Il suo primo obiettivo è stato quello di studiare gli impatti della digitalizzazione sulle principali aree di competenza dell'Istituto, dall'offerta alla commercializzazione, dalla prestazione dei servizi di investimento alla negoziazione sul mercato secondario. Considerando, ad esempio, il mondo degli intermediari finanziari, la Consob si è focalizzata sui nuovi modi di interfacciarsi con i clienti e prestare servizi di consulenza o gestione (si pensi ai Robo advisors). Questi temi non sono stati trattati in maniera isolata, ma sono stati oggetto di esame e confronto con altre autorità di vigilanza europee, all'interno di ESMA.

Tra i vari studi promossi dallo *Steering Committee Fintech* si ricorda il documento sulle *Initial Coin Offerings* e quello sulle *Security Token Offering*. Questi approfondimenti sono stati particolarmente utili anche nel processo di partecipazione alla formulazione della normativa europea, dove sono attualmente in corso di approvazione il progetto di regolamento *Markets in crypto-assets* – MICAR e il cosiddetto *pilot regime*.

In questo ambito sono state dunque analizzate alcune prime iniziative relative ai *crypto asset* e

altre iniziative proposte da intermediari o altri soggetti. Per comprendere le sfide che ne derivano è stato, inoltre, creato un incubatore per l'innovazione, come strumento di confronto tra i soggetti regolamentati e non, da un lato, e la Consob, dall'altro. Si tratta di un punto di contatto volto a promuovere un ambiente di conoscenza favorevole e utile sia ai soggetti che hanno interesse a interfacciarsi con l'Istituto, sia per la stessa Consob intenzionata a osservare l'evoluzione del mercato al fine di modificare, di conseguenza, le attività di vigilanza.

La Consob si è poi posta l'obiettivo di capire come utilizzare essa stessa le nuove tecnologie per supportare e potenziare il proprio operato. A questo riguardo, è stato adottato un approccio S.M.A.R.T. Questo acronimo è riconducibile alle seguenti parole chiave.

La digitalizzazione come "sistema", tale da coinvolgere tutte le unità operative nel processo di digitalizzazione. "Modulare": la Commissione ha infatti immaginato di sviluppare dei "moduli" che poi potessero essere, con le opportune integrazioni e modifiche, esportati o importati da altre unità operative. "Adottare": tutte le tecnologie sono al servizio dell'attività della Consob e devono essere, pertanto, flessibili così da poter essere utilizzate nel modo più efficace ed efficiente possibile. "Risolvere", nel senso che le tecnologie devono supportare la Commissione nella soluzione di eventuali problemi connessi all'attività istituzionale, migliorando le tecniche di vigilanza e le modalità di interazione con i soggetti vigilati. "Testare": qualsiasi nuova tipologia di digitalizzazione che si inserisca nel modello di vigilanza deve essere testata, dal momento che la Consob deve essere nelle condizioni di comprovarne, anche nei confronti di soggetti terzi, l'efficacia e la validità.

Fatte queste premesse generali, qual è l'organizzazione che al momento è coinvolta in questo



processo? Le singole divisioni, lo *Steering Committee* (che, come visto, funge da collettore e motore di stimolo per i settori *legal-tech*, *sup-tech* e *reg-tech*), l'*incubator* (che studia progetti di sperimentazione per capire in che misura la digitalizzazione può aiutare a meglio realizzare gli obiettivi di vigilanza e a renderli più efficaci ed efficienti), la *sandbox* presso il MEF (alla cui disciplina la Consob contribuisce insieme a Banca d'Italia e IVASS) e, ultimo nato, il *liftech*. Quest'ultimo ambiente rappresenta un canale di comunicazione volto a favorire un più agevole confronto tra la Consob e gli operatori FinTech interessati ad accedere alle sperimentazioni, per capire se e in che misura i loro progetti sono di interesse e possano rientrare nella sfera di competenze dell'Istituto.

1.1. La vigilanza sui prodotti finanziari

Alessandra Limosani

Responsabile Ufficio prodotti CONSOB

Vi è la tendenza a pensare all'intelligenza artificiale come a un algoritmo complicato, capace di elaborare grandi masse di dati ed effettuare ipotesi predittive, estraendo conoscenza dai dati. L'algoritmo effettua le analisi direttamente sui dati (*data driven*) e non è spiegabile,

Esiste tuttavia anche un altro tipo di intelligenza artificiale, finalizzata alla costruzione di *dataset* di elevata qualità, che serve a predisporre l'ecosistema dei dati sui quali applicare *analytics* di tipo avanzato. Questa attività di “preparazione” (*data preparation*) è il presupposto dei *data analytics*, se si pensa che ben l'80% del tempo dei *data scientist* è dedicato a raccogliere, collazionare, ma soprattutto a “pulire” i dati (“eliminare il rumore”), rendendoli utilizzabili e fruibili all'applicazione dei *tool* di *advanced analytics*.

In sintesi, i due approcci danno luogo alla costruzione di due sistemi di intelligenza artificiale diversi. Il primo è quello *data driven*, in cui l'algoritmo cerca di estrarre la conoscenza dall'analisi dei dati; il secondo, *model driven*, si basa sulle regole definite dagli esseri umani e la macchina svolge il compito assegnato sulla base dei criteri predefiniti.

Questi due sistemi, molto studiati nei decenni scorsi, sono stati in passato considerati come quasi contrapposti. Ultimamente si è compreso che la loro integrazione dà valore aggiunto alla possibilità di usare le nuove tecnologie, consentendo anche di controllarle, di ottenere la spiegabilità dei modelli e di accrescere l'accuratezza e la precisione dei risultati.

Una prima sperimentazione è stata applicata a supporto dell'attività di vigilanza sul KID (*key information document*) dei PRIIPs, ossia il documento sintetico che illustra le caratteristiche

chiave dei prodotti finanziari PRIIPs (*Packaged Retail and Insurance-based Investment Products*), prodotti pre-assemblati rivolti agli investitori al dettaglio il cui *payoff* è complesso, perché dipende dall'andamento di variabili sottostanti. Il KID dei PRIIPs è prescritto dal Regolamento europeo n. 1286/2014 ed è pensato come un documento cartaceo in formato pdf e *non machine readable*. In altre parole, il legislatore europeo non ha fatto lo sforzo di pensare già a una “regolamentazione digitalizzata”.

Dal punto di vista della vigilanza, il ricorso alle tecnologie basate sull'uso dell'intelligenza artificiale assolve all'importante funzione di accrescere l'efficienza delle selezioni e delle analisi di tipo campionario. L'elevata numerosità dei documenti in questione, infatti, rende impossibile uno scrutinio completo di tutti i documenti, e richiede la definizione di criteri di selezione evoluti e non puramente casuali, in modo da ridurre il rischio di vigilanza. Si pensi che solo nel 2018 sono stati raccolti circa un milione di documenti e il *trend* è risultato in forte crescita negli anni successivi.

In quest'ottica sono stati raccolti un ristretto set di dati strutturati in modo da filtrare questi documenti e individuare alcuni indicatori sintetici che permettessero di selezionare i documenti da “attenzione”.

L'obiettivo della sperimentazione è stato quello di automatizzare il processo di estrazione di informazioni dai documenti e definire una “ontologia” che collegasse i concetti astratti mettendoli in relazione, giungendo a elaborare una tassonomia dei prodotti finanziari. Si è pervenuti così a clusterizzare i diversi tipi di prodotti in maniera automatizzata, raggruppandoli secondo caratteristiche omogenee. La Consob ha definito le regole del *mapping* e chiesto l'estrazione dei campi secondo le regole da essa stessa definite, al fine di ottenere risultati precisi e spiegabili.



Nel 2019 è stato sviluppato il primo *proof of concept* (PoC), in *partnership* con un *team* di esperti del dipartimento di ingegneria informatica dell'università La Sapienza di Roma (selezionato in base all'*h-index*). Il prototipo si basa sull'applicazione di una serie di tecniche di analisi del linguaggio naturale e mira all'estrazione di parole chiave e di concetti che si trovano in uno specifico campo dentro il documento.

Tra i documenti relativi al 2018 è stato selezionato un campione oggetto dei primi *test*, per poi arrivare all'estrazione di alcuni campi. Successivamente, sono stati estratti da questi documenti i concetti di interesse; infine, i documenti sono stati clusterizzati in base alle tipologie di prodotti.

Questo prototipo è stato ulteriormente affinato: il nuovo *test* condotto su 17.000 documenti che si riferiscono a 182 ideatori suddivisi in 27 classi di prodotti ha dimostrato un grado di precisione e accuratezza elevatissimo, pari al 99% (un risultato importante, se si considera che alcuni modelli algoritmici vengono considerati "buoni" pur avendo dei tassi di accuratezza al 60%).

In questo momento è allo studio l'applicazione di una tecnologia di *machine learning* per cercare di far evolvere le regole (disegnate sullo storico) in modo dinamico, così da adattarsi al cambiamento dei prodotti, dell'offerta e dei mercati.

L'obiettivo di lungo periodo (quasi utopistico) è ideare un algoritmo che consenta automaticamente di indicare le priorità delle azioni di vigilanza, che sia in grado di effettuare uno *screening* automatico e di selezionare direttamente i documenti su cui focalizzare il lavoro dell'analista.

1.2. La vigilanza sugli abusi di mercato

Giuseppe Frega

Coordinatore Ufficio vigilanza fenomeni abusivi CONSOB

La Consob sta conducendo approfondimenti per sviluppare l'applicazione di strumenti informatici all'attività di contrasto agli abusivismi finanziari. Le attività finanziarie abusive sono poste in essere per il 75% *online* e si caratterizzano per essere particolarmente invasive proprio grazie all'utilizzo di tecnologie digitali che hanno, da un lato, sviluppato forme di rapida interazione a distanza e, dall'altro, consentito la realizzazione di nuove modalità operative.

Pertanto, la maggior parte delle imprese che presta abusivamente attività finanziarie è in grado di superare agevolmente i confini nazionali ed operare *cross border* mediante canali telematici e digitali quali siti *web* interattivi, *social media* e infrastrutture tecnologiche che si basano su registri elettronici distribuiti (DLT), interfacciandosi con un significativo numero di clienti di differenti nazionalità. Tali imprese si rivolgono alla clientela facendo non di rado leva su (vere o presunte) innovazioni di processo o di prodotto, ponendosi in sleale concorrenza con i soggetti abilitati, sviando una quantità non irrilevante di risparmio e impattando, quindi, su un'efficiente allocazione delle risorse in ambito finanziario.

Nell'ambito del piano di digitalizzazione della Consob di cui al paragrafo 1, è in fase di sviluppo un progetto di *intelligent crawling* che ha l'obiettivo di dotare la Commissione di un motore di ricerca che, in futuro, anche con capacità crescente di autoapprendimento, dovrà scandagliare il *web* per intercettare sempre più velocemente le piattaforme *online* mediante le quali sono prestate abusivamente attività riservate, riducendo le operazioni manuali e incrementando

il numero delle azioni a contrasto del crescente fenomeno degli abusivismi finanziari.

Il prototipo di questo *software*, dovrà essere testato per valutarne poi appieno le possibilità di effettivo utilizzo. Il progetto informatico della Consob per il contrasto alle attività abusive si sviluppa complessivamente su due direttrici: la realizzazione, da un lato, di un applicativo per l'analisi semantica degli esposti e, dall'altro - come accennato - la realizzazione di un *software* per l'individuazione e l'analisi dei siti *web* mediante i quali sono poste in essere abusivamente attività finanziarie riservate.

Il sistema di analisi automatica degli esposti è teso a individuare i dati di maggiore interesse presenti negli esposti in tema di abusivismo che pervengono alla Consob. Tali dati vanno poi ad alimentare il *software* di *crawling* che li processa per l'individuazione e l'analisi dei siti *web* mediante i quali sono prestate le attività abusive. Il processo informatizzato si avvia quindi con l'analisi semantica degli esposti che si articola principalmente nelle fasi di “*keyword analysis*” e “*topic analysis*” offrendo anche funzioni di “mappa semantica” e “analisi delle caratteristiche dell'autore dell'esposto”.

La *keywords analysis* è una funzionalità che esamina un *set* di parole ricorrenti, individuate anche in base ad un elenco di parole sensibili previamente selezionate, che poi si andrà ampliando man mano che il sistema acquisirà la terminologia presente negli esposti. Questa è una funzione di intelligenza artificiale che processa il linguaggio naturale individuando le parole rilevanti, organizzandole per renderle fruibili per il successivo *step* di *topic analysis*.

La *topic analysis*, attraverso la funzione di *text mining*, utilizza l'elaborazione del linguaggio naturale per trasformare il testo libero di documenti in dati strutturati e normalizzati. Questa è una funzione di intelligenza artificiale che

processa, quindi, il linguaggio naturale collegando le parole (“*keywords*”) precedentemente individuate, evidenziando gli argomenti rilevanti. La *topic analysis* offre una visione interattiva ad alto livello degli argomenti più significativi in base alle parole più ricorrenti nell’esposto.

La funzionalità del sistema di analisi degli esposti denominata “mappa semantica” collega *keywords* e *topic*, in modo da offrire una visione d’insieme delle tematiche presenti negli esposti, permettendo così di operare collegamenti che consentono lo svolgimento di analisi a più ampio spettro: si tratta di analisi di tipo aggregato che consentono di fare ricerche in base a elementi che presentano analogie.

Infine, la funzione di “analisi delle caratteristiche dell’autore dell’esposto” consente di valorizzare informazioni di contesto su quest’ultimo quali categoria professionale, area geografica etc.

Il processo informatizzato di analisi degli esposti permette, nel complesso, di creare una rete di collegamenti tra i soggetti menzionati negli esposti al fine di individuare eventuali *network* di persone, società, località geografiche che facciano emergere azioni concorrenti nella commissione di illeciti di abusivismo.

A partire dalla base di informazioni ricevute dall’analisi degli esposti si attiva la funzione del *crawler* che, in prospettiva, dovrebbe svolgere (secondo modalità automatiche) accertamenti *online* su tutti i soggetti e indirizzi *web* menzionati negli esposti, anche al fine di individuare possibili collegamenti fra gli stessi e verificare poi gli eventuali siti *web* con indirizzo affine a quelli indicati negli esposti. Il *crawler*, quindi, registra le informazioni individuate, raccolte *online*, effettuando il *download* delle pagine rilevanti dei siti *web* esaminati in cui vi sono elementi sull’operatività illecita; registra la lingua in cui il sito è disponibile, nonché le informazioni disponibili su questi siti, sull’identità di soggetti a cui risulta

riconducibile l’attività illecita ed eventualmente la localizzazione geografica degli stessi. Nello svolgimento di tali accertamenti, il *crawler* verifica in automatico se vi siano o meno *warnings* da parte di altre Autorità. Al termine della prima fase di sviluppo del progetto, attualmente in corso, il *crawler*, registrate le informazioni contenute nei siti *web* esaminati, dovrebbe quindi procedere alla creazione di un *report*, vale a dire una scheda informativa che fornisca al funzionario un quadro sinottico dei contenuti di ciascun sito *web* esaminato, agevolando anche la materiale predisposizione di provvedimenti e documenti ufficiali della Consob, attraverso l’estrazione di informazioni dai *report* automatizzati. La definizione del *report* costituisce la parte più delicata che dovrebbe altresì costituire la base per un processo di *machine learning* che dovrebbe portare il *software* ad effettuare in autonomia (non solo sulla base delle informazioni rivenienti dagli esposti) una *detection* dei siti *web* mediante i quali sono svolte attività abusive con classificazione automatica degli stessi.

Quest’ultimo *step*, costituito da una fase più avanzata dell’automatizzazione dell’attività di *detection* dei siti *web* abusivi sarà eventualmente sviluppato a valle della prima fase di realizzazione del progetto informatico, tuttora in corso, a supporto dell’attività di contrasto agli abusivismi finanziari.

1.3. La vigilanza sui mercati

Paola Deriu


Responsabile Ufficio vigilanza operatività mercati a pronti e derivati CONSOB

Nell’ambito delle proprie attività di vigilanza sui mercati la Consob deve garantire l’ordinato svolgimento delle negoziazioni, nonché prevenire, individuare e sanzionare possibili fenomeni di abusi di mercato. Si tratta di un compito in divenire, reso sempre più difficile dalla crescente mole di informazioni e dati da acquisire, elaborare ed interpretare. Tali dati, strutturati, riguardano sia transazioni (in termini di contratti conclusi tra due controparti sui diversi mercati regolamentati e sulle altre piattaforme di scambio sottoposte alla vigilanza della Consob), sia proposte di negoziazione (ordini) che possono aver dato luogo alla conclusione di transazioni o essere stati modificati, cancellati, revocati, eseguiti solo parzialmente. La crescita del numero e della dimensione delle piattaforme di negoziazione, degli strumenti finanziari negoziati, della platea di intermediari ed investitori ha notevolmente ampliato il perimetro da vigilare ed incrementato gli *input* di vigilanza da prendere in considerazione (reclami e segnalazioni di operazioni sospette di costituire abusi di mercato, segnalazioni da parte di altri soggetti pubblici); di conseguenza i *data set* da analizzare continuano a crescere in maniera esponenziale.

In tale ambito, già da tempo la Consob si avvale di un sistema di “allertatori intelligenti”, basati su analisi statistiche ed algoritmi, in sostanza su un concetto di intelligenza artificiale intesa in senso lato; si tratta di circa 200 *alert* che segnalano anomalie che si verificano sulle *trading venue* vigilate. Il grado di complessità delle analisi sottostante a ciascun *alert* è variabile; i casi più complessi riguardano *pattern* e scenari elaborati utilizzando più algoritmi concatenati sulla base di

logiche e procedure di analisi definite nel tempo sulla base delle esperienze di vigilanza, anche grazie al confronto con omologhe strutture di vigilanza di autorità estere. Gli *alert* segnalano le condotte operative da approfondire.

L’obiettivo, condiviso in Europa e nel mondo, è che si possano definire sistemi intelligenti in grado di ottimizzare le analisi e gli approfondimenti che vengono condotti a valle dei segnali di anomalie; ottimizzazione che deve tener conto anche dell’incremento dei dati da analizzare e conseguentemente degli *input* di vigilanza. Il lavoro di analisi e “scrematura” che viene condotto per ciascun *alert* è frutto di una serie di elaborazioni e ragionamenti effettuati dagli analisti e validati sulla base di una serie di regole e di prassi introiettate dall’Istituto: non è detto che una macchina sia in grado di replicare lo stesso tipo di “intelligenza” che applica un analista esperto nella valutazione dei casi concreti. Al riguardo, nel rammentare che uno degli obiettivi dei modelli di supervisione sull’ordinato svolgimento delle negoziazioni riguarda l’individuazione di possibili condotte potenzialmente riconducibili ad abusi di mercato, si evidenziano importanti differenze tra *insider trading* e manipolazione. I modelli per la *detection* di condotte di *insider trading* sono relativamente semplici da costruire, in quanto incardinati sulla pubblica diffusione di un’informazione *price sensitive* che provoca un impatto sui prezzi e sui volumi sulla base del quale impostare: a) una analisi retrospettiva del *trading* antecedente all’evento, b) una ricerca di eventuali collegamenti tra posizioni direzionali rilevate con il circuito di formazione dell’informazione privilegiata, c) una valutazione circa la (dis)continuità nella *trading history* di ciascuna posizione analizzata. Diversamente, i modelli per la *detection* di casi potenzialmente riconducibili a manipolazione del mercato sono molto più articolati e complessi: al


 riguardo si consideri che anche a livello normativo i molteplici schemi manipolativi individuati dal legislatore non rappresentano una casistica esaustiva. A titolo esemplificativo, operando una notevole semplificazione per ragioni di brevità, schemi manipolativi possono essere realizzati sia con attività di *trading* (*trade-based manipulation*), sia tramite la diffusione di informazioni (*information-based manipulation*), tra le quali la diffusione di *fake news*, ovvero tramite altri artifici (*action-based manipulation*), cosicché la definizione delle architetture di modelli che puntano ad individuare ed analizzare segnali di possibili condotte di manipolazione risulta molto più complessa, granulare e modulare.

Come detto, il sistema di vigilanza sui mercati adottato dalla Consob fa già uso di algoritmi intelligenti. Tra i vari *pattern* che riguardano la manipolazione, il sistema è in grado di individuare, ad esempio, condotte di *pump and dump* nel caso in cui in una determinata seduta un investitore abbia cercato di dare un *imprinting* in crescita (*pump*) al prezzo di un titolo, così da riuscire a vendere (*dump*) grosse quantità a prezzi che, in assenza del *pump*, non sarebbe stato possibile ottenere. Attualmente tali segnali vengono esaminati da un analista al fine di valutarne la gravità e l'opportunità di approfondimento ed *escalation*; in particolare l'analista effettua una serie di elaborazioni ulteriori, volte a valutare se e quanto l'operatività dell'investitore sia stata "aggressiva" o "subita", se può essere il frutto di una coincidenza, se i contratti sono stati conclusi tra controparti ricorrenti, ecc. L'obiettivo che si intende perseguire in questa fase, in relazione al quale la Consob ha avviato un progetto con il dipartimento di matematica di una primaria istituzione accademica italiana, consiste nel definire modelli di *supervised machine learning* in grado di replicare i ragionamenti che solitamente vengono fatti dagli analisti. Tale progetto è in fase

più embrionale rispetto ai progetti definiti ai paragrafi 1.1 e 1.2 e seguirà un approccio articolato in più moduli.

Un primo modulo riguarda lo sviluppo di metodi statistici che favoriscano le analisi di *big data* in maniera standardizzata, attraverso funzioni di clusterizzazione degli investitori in relazione sia alla loro *trading history* (intensità, *magnitudo*, frequenza, *asset class*, aggressività, profittabilità, etc.) sia, nello specifico, a eventi *price sensitive*. Il secondo modulo riguarda l'ottimizzazione dei segnali di allarme; ad esso sarà collegato il terzo modulo che verterà sullo sviluppo di segnali predittivi di fattispecie manipolative o di condotte scorrette con particolare attenzione al tema dell'amplificazione e della propagazione di piccoli *shock* che colpiscono il sistema di mercato. La quarta parte riguarderà lo sviluppo di sistemi idonei all'utilizzo ed all'integrazione dei dati da fonti alternative, ad esempio dai *social network*, al fine di identificare schemi di manipolazione dei prezzi coordinati anche tramite l'utilizzo di piattaforme non tradizionalmente poste sotto la vigilanza dei regolatori.

Su questo ultimo punto va dato atto dell'intenso dibattito in corso, di carattere politico, accademico e regolatorio, scaturito da alcuni fatti che si sono verificati nel gennaio 2021, principalmente sui mercati statunitensi. Si fa riferimento alla cosiddetta "vicenda GameStop": nell'arco di poche sedute, comprese tra il 12 ed il 27 gennaio 2021 il prezzo delle azioni della società, quotata al NYSE è passato da \$ 20 a \$ 483 (massimo *intra-day* del 27 gennaio 2021) per poi flettere nuovamente e tornare ad oscillare nella banda di \$ 40-50, nel mese di febbraio. La significativa volatilità dei prezzi delle azioni GameStop (così come di altre azioni, cosiddette "*meme stock*", sebbene con minori escursioni), accompagnata da un volume di scambi esponenzialmente cresciuto è il risultato combinato di una serie di fattori, in

parte tecnici, in parte riconducibili al ruolo del *social media*, attraverso i quali è stato realizzato un vero e proprio *flash mob* finanziario. Gli incrementi del prezzo delle azioni GameStop sono stati determinati da una miriade di acquisti effettuati da giovani investitori *retail* che hanno risposto a una sorta di “chiamata alle armi” contro “i potenti” di Wall Street, attivi sui mercati finanziari tramite piattaforme accessibili attraverso *app* gratuite e di facile utilizzo. Tali investitori sono stati esortati prima all’acquisto e poi al mantenimento delle posizioni tramite canali *social*, in particolare il *forum* “WallStreetBets” della piattaforma Reddit. L’interesse per l’azione da parte della comunità *on line* non si basava tanto su valutazioni societarie, né appariva prevalentemente finalizzato al mero conseguimento di obiettivi finanziari; esso veniva ricondotto, almeno in una prima fase, al desiderio di una risposta “dal basso” a quella che veniva descritta come una speculazione del cosiddetto *establishment* di Wall Street, nella fattispecie da *hedge fund* speculativi che, ben prima dell’avvio del massiccio *trend rialzista*, avevano assunto posizione nette corte, scommettendo in sostanza sulla diminuzione del prezzo delle azioni. L’ampia partecipazione di massa al *momentum* è stato possibile anche grazie all’utilizzo di piattaforme di *trading* cosiddette *zero commission* che, oltre a dichiarare di non applicare alcuna commissione consentono anche investimenti di controvalori molto bassi. L’intera vicenda, tuttora sotto esame da parte delle competenti autorità statunitensi, così come da parte di molteplici autorità di altri Paesi, costituisce un punto di svolta nelle analisi dell’intersezione tra tecnologia e finanza. I temi sui quali il dibattito è molto intenso sono diversi e tutti meritevoli di riflessioni (solo per menzionarne alcuni si tratta della *gamification* del *trading* tra democratizzazione e necessità di tutela degli investitori, dell’adeguatezza delle

regole che presidiano lo svolgimento delle negoziazioni, della tutela degli interessi dei clienti in relazione a pratiche di PFOF - *payment for order flow*, dell’efficacia delle strategie di comunicazione da parte delle autorità di vigilanza rispetto a un pubblico di investitori più avvezzo ai *tweet* che ai comunicati ufficiali, di nuove (?) forme di manipolazione del mercato.

In questa sede pare opportuno sottolineare che, nell’ambito dei progetti in via di sviluppo, l’analisi incrociata tra dati di vigilanza e dati provenienti da fonti aperte, tradizionalmente non riconducibili a entità sotto la diretta supervisione delle autorità di vigilanza, costituisce una importante sfida da cogliere.

*W&L
Law*

2. AGCOM e l'intelligenza artificiale per fronteggiare l'hate speech

Gennaro Ragucci

Funzionario Servizio economico-statistico AGCOM

L'Autorità per le garanzie nelle comunicazioni può essere considerata sia un utilizzatore, che un produttore di dati. In particolare, raccoglie dati dagli operatori delle comunicazioni, fa parte del sistema statistico nazionale (SISTAN) e fornisce indicatori e statistiche a vari organismi nazionali e internazionali. L'Autorità è dotata di un'apposita struttura a questo preposta che inoltre, anche in collaborazione con le altre strutture, svolge analisi, effettua indagini e produce osservatori periodici. Inoltre, al pari della Consob, i *data scientist* che lavorano al suo interno si avvalgono anche della collaborazione con università e centri di ricerca per lo sviluppo di strumenti di intelligenza artificiale.

Per l'autorità, l'intelligenza artificiale costituisce uno strumento per l'analisi avanzata dei dati (cosiddetta *AI analytics*) e anche un potenziale oggetto di regolazione (si veda a tal proposito anche la recente proposta di regolamento europeo sulla materia). È infatti necessario comprendere i fenomeni che coinvolgono i *big data*, per fare un'analisi del valore dei dati in possesso degli operatori delle comunicazioni e delle relazioni articolate che si instaurano tra piattaforme/servizi *online* e consumatori.

Per quanto concerne il formato dei dati, l'Autorità tratta sia dati strutturati, che sono forniti essenzialmente dagli operatori della comunicazione, ad esempio dati economici e infrastrutturali, sia dati semi strutturati e non strutturati, essenzialmente dati testuali, come gli articoli pubblicati sui media tradizionali e sui *social*. L'intelligenza artificiale costituisce un valido supporto soprattutto nel caso di dati non strutturati.

Quanto al governo dell'intelligenza artificiale, occorre chiedersi se, e quando sia opportuno aprire la cosiddetta "scatola nera". Dal punto di vista dell'Autorità ciò è importante quando l'IA impatta su ambiti di competenza istituzionale, come la tutela dei consumatori, della concorrenza sui mercati e il pluralismo informativo. L'Agcom tende alla massima apertura e *disclosure* sui dati e degli algoritmi. Tuttavia, il problema resta aperto ed è attualmente oggetto di ricerca: non è facile aprire questa scatola nera e monitorare di conseguenza la correttezza dei risultati, perché vi sono problematiche, lato operatore, legate al fatto che, quando si esamina un componente di intelligenza artificiale i dati possono essere coperti da *privacy* o segreto commerciale, così come possono esserlo gli algoritmi. Inoltre l'apprendimento automatico (*machine learning, deep learning*) si basa su tecniche "non esplicite" di programmazione dei computer e questo complica in generale l'interpretazione dei risultati.

Un altro ambito di applicazione dell'intelligenza artificiale in autorità è costituito dalla tematica dell'*hate speech*. A questo riguardo, l'AGCOM partecipa a un progetto della Commissione europea, IMSyPP "Innovative Monitoring Systems and Prevention Policies of Online Hate Speech". Inoltre ha realizzato studi e analisi soprattutto sui temi della qualità dell'informazione, ad esempio, attraverso tecniche di NLP (*Natural Language Processing*), che consentono di estrarre conoscenza ad esempio da *fake news*.

Per spiegare meglio questa attività, può essere utile analizzare un caso concreto. L'autorità si è dotata di un servizio che consente di raccogliere tutti i documenti di testo relativi alle diverse notizie pubblicate. Ad esempio, utilizzando questo servizio, in una cartella sono state raccolte tutte le notizie del mese di aprile 2020, che possono arrivare ad alcune migliaia di articoli di giornale.

Per estrarre informazioni utili riguardo ai temi trattati da questa vasta mole di dati, è possibile utilizzare tecniche di modellazione degli argomenti (*topic modeling*), ossia un modello statistico per l'individuazione automatica degli argomenti latenti che compaiono nella raccolta documentale. Per effettuare un riconoscimento automatico dei *topic* occorre realizzare una *data pipeline*, costituita, oltre che alla anzidetta raccolta preliminare dei testi delle notizie, da una fase di preparazione dei dati (di *tokenizzazione*, vale a dire separazione delle frasi in singoli termini) e una fase di “eliminazione del rumore” (che può consistere nella eliminazione delle *stop words*, come le congiunzioni “e” e “o”). Questi termini poi vengono “preparati” attraverso la loro trasformazione in numeri e in vettori, perché gli algoritmi poi di fatto possono operare solo su dati numerici. In questo caso è stato utilizzato un algoritmo specifico detto LDA (*Latent Dirichlet Allocation*). Per la visualizzazione dei risultati ottenuti è stato infine utilizzato lo strumento “LDAvis”. L'Autorità ha così realizzato un'analisi dei quotidiani dell'aprile 2020 e ha potuto osservare, come era facile aspettarsi, che i termini più aggregati che fanno parte poi dei *topic*, automaticamente individuati, hanno a che fare maggiormente con le tematiche del coronavirus, degli ospedali e così via. I risultati di analisi similari a questa sono confluiti nell' “Osservatorio sulla disinformazione *online*” realizzato dall'autorità.



3. ARERA e l'“intelligenza assistita”, l'“intelligenza condivisa” e l'“intelligenza della filiera”

Luca Lo Schiavo

Vice direttore Direzione Infrastrutture, Energia e Unbundling ARERA

Luca Lazza

Responsabile Unità servizi informativi ARERA

ARERA non fa uso diretto di applicazioni di intelligenza artificiale. Tuttavia, l'analisi dei dati costituisce un elemento fondamentale per la regolazione e nel tempo si sono sviluppate esperienze sempre più evolute, che possono essere classificate su tre livelli progressivi di utilizzo di intelligenza nell'analisi dei dati.

3.1. “Intelligenza assistita”: l'esperienza con SAS Visual Analytics

L'autorità elabora grandi moli di dati, in contesti anche molto differenti tra loro. A fronte di una regolazione che si può considerare “stabilizzata” in quanto riferita a mercati maturi e oggetto di regolazione da molti anni, nell'ambito della quale vengono effettuati il calcolo delle tariffe e supportato il processo di vigilanza e monitoraggio, vi sono attività del regolatore legate ai nuovi ambiti oggetto di regolazione, sia per quanto riguarda nuovi settori affidati per legge alla regolazione (come il settore dei rifiuti), sia per quanto concerne nuovi sviluppi nei settori già regolati da tempo (come i servizi di flessibilità nel settore elettrico). Questi comportano una gestione della regolazione diversa e per certi aspetti più complessa, in ampia misura legata a una logica di tipo sperimentale, intrinsecamente correlata alle nuove attività o alla novità della regolazione. Dovendo coniugare le esigenze di analisi “stabilizzate” con quelle “sperimentali”, l'autorità si è dotata degli strumenti SAS Visual Analytics e Visual Statistics. Si tratta di un sistema di

reportistica e analisi statistica per l'esplorazione visuale e intuitiva di una grande mole di dati. Attraverso questi strumenti si rappresentano i dati secondo diverse modalità ed è possibile cambiare gli assi di riferimento di questi dati, gli intervalli temporali, gli intervalli spaziali, gli intervalli di posizioni geografiche in maniera intuitiva, permettendo così ai data scientist e ai referenti degli uffici tecnici di esplorare il dato da diversi punti di vista, aggregando o separando gli elementi di base che lo compongono, agevolando così l'individuazione di trend che prima risultavano difficili da vedere. Le analisi svolte in mancanza di questi strumenti richiedevano personale dotato sia di competenze specifiche nel settore di competenza, che conoscenze informatiche evolute. Oggi non è più così: chiunque sappia usare un browser e conosca il settore può elaborare questo tipo di analisi. Inoltre, gli strumenti SAS permettono di semplificare moltissimo la parte di “miglioramento” della qualità del dato (ad esempio, segnalando quando un dato è poco strutturato o “sporco” o si combinano dati provenienti da fonti diverse) e quindi riducono anche moltissimo il tempo necessario per passare dal dato grezzo al dato ragionato e ripulito. Per questo motivo, possiamo parlare di strumenti di “intelligenza assistita”.

Gli strumenti SAS sono stati utilizzati per la prima volta con riferimento alla revisione della regolazione della qualità dei servizi, che viene effettuata con cadenza quadriennale in virtù di una *sunset clause* prevista dalla normativa. In particolare, nel passaggio dal quadriennio 2016-2019 al successivo è stata compiuta una valutazione di impatto della regolazione-VIR (documento di consultazione n. 287/2019) con un'ampia analisi quantitativa che contiene alcune “viste” innovative dei dati di continuità della fornitura elettrica che sono risultate propedeutiche a immaginare soluzioni regolatorie nuove ad alcuni problemi. A



titolo esemplificativo, uno dei temi è quello delle grandi città con riferimento alle quali gli strumenti di SAS hanno consentito di mettere in evidenza il margine di errore nell'attribuire a una certa rete una determinata concentrazione territoriale.

In generale, si può dunque dire che spesso i sistemi di “intelligenza assistita” favoriti dalla tecnologia di analisi visuale dei dati supportano la produzione dei provvedimenti dell'autorità e l'impostazione delle consultazioni e delle VIR.

3.2. “Intelligenza condivisa”: l'esperienza REMIT (con ACER)

Il caso che potrebbe essere definito di “intelligenza condivisa” consiste in un trattamento di dati di *trading* dell'energia secondo il regolamento europeo REMIT-*Regulation on energy market integrity and transparency*, volto al monitoraggio delle transazioni sui mercati all'ingrosso dell'energia e all'individuazione automatica di determinati eventi che potrebbero nascondere casi di abuso di mercato. L'Agenzia europea di collaborazione dei regolatori dell'energia-ACER, ARERA, in collaborazione con le altre autorità nazionali di regolazione dell'energia, ha sviluppato uno strumento per il monitoraggio dei mercati energetici all'ingrosso che ha l'obiettivo principale di verificare il corretto operato degli operatori coinvolti nel mercato ed individuare attività abusive (con un effetto che è anche di carattere preventivo-dissuasivo).

Si tratta di vigilanza sui casi di *market abuse*, già ampiamente esercitata dalle autorità finanziarie (come Consob) sui mercati mobiliari: vi sono però delle rilevanti differenze tra il caso di REMIT e quello che avviene nei mercati finanziari “tradizionali”. Innanzitutto, REMIT tiene conto anche di tutta una serie di informazioni legate a dati fondamentali: non soltanto i prezzi acquisto-ven- dita, ma anche la capacità delle unità di

generazione, lo stato delle interconnessioni e altri dati che influenzano i consumi (ad esempio di carattere calendariale e meteorologico); inoltre, tiene conto di tutti i possibili mercati di *trading* all'ingrosso di energia elettrica nell'ambito dell'Unione, di tutte le transazioni e di tutte le tipologie di prodotti, fra cui sia gli acquisti interni ad un singolo Paese, sia le transazioni *crossborder*. Il sistema di monitoraggio REMIT si pone, dunque, l'obiettivo di riuscire a tenere “sotto controllo” tutto ciò che riguarda i mercati energetici all'ingrosso, in ogni loro declinazione e tenendo conto di tutti i fondamentali dati che sono utilizzati, oltre ai prezzi e alle quantità scambiate. Per farlo utilizza uno schema che si basa, al primo step, sull'acquisizione di dati. Il core di REMIT consiste poi, sulla base di tutte le transazioni e di tutti gli eventi analizzati, nella individuazione di pattern di gruppi di transazioni, per identificare gli eventi anomali, cioè le transazioni che si discostano in maniera significativa (alla luce di tutto l'insieme dei parametri) da quella che è la norma.

A questo punto, dal processo di monitoraggio automatico si apre un sotto processo di verifica manuale, organizzata in due fasi. La prima analisi preliminare, svolta da ACER, è legata alla validazione di ciò che lo strumento automatico ha fatto, per capire se effettivamente si può considerare un evento come sospetto, oppure ciò può essere frutto di un errore del modello.

Una volta qualificati gli eventi anomali come sospetti, la seconda fase si svolge all'interno di ARERA (e delle altre autorità nazionali di regolazione in ambito euro-unitario) ed è volta all'istruttoria sui casi sospetti individuati e “validati”, tenendo conto delle eventuali informazioni aggiuntive a disposizione dell'autorità ed effettuando appositi approfondimenti ulteriori. Vi è poi un *feedback loop* che, sulla base delle risultanze dell'analisi e di quanto viene segnalato



dai tecnici, porta ad aggiornare il modello su base mensile.

Questo sistema consente una riduzione del 90% dei casi che richiedono l'analisi da parte di un operatore dell'autorità. Un aspetto distintivo di questo sistema è rappresentato dal collegamento con i *fundamental data* (come la capacità di generazione, lo stato delle linee di interconnessione, la capacità di transito), che sono nel settore elettrico particolarmente variabili (per esempio in relazione alle fluttuazioni orarie dei consumi, delle immissioni e quindi dei prezzi). Un altro aspetto significativo è la centralizzazione di questo controllo, che permette di avere a disposizione moli di dati impressionanti, oltrepassando la dimensione nazionale e abbracciando il livello unionale nella sua interezza. Peraltro, il sistema permette all'agenzia europea di avere un quadro completo di ciò che succede in tutta Europa, risultato cruciale in quanto queste forme di manipolazione potrebbero giocarsi tra piazze diverse con riflesso su altre operazioni.

3.3. "Intelligenza della filiera": il caso smart metering 2G

La seconda generazione di smart meter di energia elettrica (la cui diffusione ha raggiunto, nel 2021, il 50% dell'utenza) è in grado di inviare direttamente e con un flusso giornaliero i dati di consumo. Inoltre, il livello di dettaglio temporale con il quale sono raccolti e resi disponibili i dati (granularità) è cambiato in modo significativo.

Se il vecchio misuratore tradizionale permetteva sostanzialmente un'unica lettura, non differenziata nel tempo e tipicamente svolta su base annua, dal 2001 (con lo *smart meter* di prima generazione) siamo passati ad avere una lettura al mese con altissima regolarità. Questa lettura al mese è mediata su tre fasce orarie, quindi vi sono tre dati al mese, contro un dato all'anno rilevato manualmente dai vecchi contatori; tuttavia

questi dati mensili arrivano con un ritardo di circa tre settimane al venditore, che è il soggetto che poi "compono" la bolletta.

Con lo *smart meter* di seconda generazione, i dati vengono raccolti ogni giorno con granularità al quarto d'ora, che è l'unità minima per i servizi di dispacciamento (che assicurano l'equilibrio tra domanda e offerta e sono ora aperti a servizi di flessibilità anche per "risorse" di piccola dimensione). Con la seconda generazione di contatori, la velocità di messa a disposizione dei dati per l'utilizzo (latenza) è stata fissata dall'autorità a 24 ore successive alla mezzanotte, entro le quali deve essere raccolta una percentuale di dati superiore al 95%. Inoltre, attraverso applicazioni locali connesse al contatore di seconda generazione (anche tramite mobile), si può arrivare alla comunicazione dei dati all'utente in tempo reale anche in termini di potenza istantanea e non solo di consumi di energia (nel quarto d'ora).

I dati che vengono trasmessi hanno una importanza cruciale sia per attivare azioni virtuose di risparmio energetico da parte di consumatori "attivi", sia per supportare la concorrenza. In particolare, l'elevata granularità dei dati consentita dalla regolazione e dal connesso passaggio ai contatori di nuova generazione, ha messo a disposizione dei venditori una grandissima base di dati che ha stimolato lo sviluppo di iniziative innovative a potenziale vantaggio dei consumatori. Vi è da dire che nella regolazione esistono dei costi e delle penalità di sbilanciamento che pagano i venditori, in relazione alla differenza fra il prelievo concreto che si è verificato in un certo giorno e il prelievo previsto (stimato dal venditore sulla base dei dati di misura disponibili). Siccome queste differenze insistono poi sul prezzo finale offerto ai clienti, la riduzione degli sbilanciamenti permette al venditore di essere più competitivo. Questo può avvenire attraverso una migliore previsione: disporre di dati per tutti

i punti di dati al quarto d'ora consente ai venditori di effettuare previsioni più accurate se sono in grado di analizzare una grandissima mole di dati, che devono essere interpretati con modelli adeguati. Esperienze concrete hanno mostrato che è stato possibile, attraverso algoritmi di intelligenza artificiale, raggiungere per alcuni venditore una riduzione di costi di sbilanciamento fino al 30%. Si tratta di risparmi che possono andare a vantaggio del cliente finale o quantomeno dei clienti che scelgono venditori che hanno questo tipo di *capability*.

*W&L
Law*

