

**COMMISSIONE VI  
FINANZE**

**RESOCONTO STENOGRAFICO**

**INDAGINE CONOSCITIVA**

9.

**SEDUTA DI MERCOLEDÌ 25 OTTOBRE 2017**

PRESIDENZA DEL PRESIDENTE MAURIZIO BERNARDO

**INDICE**

	PAG.		PAG.
<b>Sulla pubblicità dei lavori:</b>			
Bernardo Maurizio, <i>Presidente</i> .....	3	Barbanti Sebastiano (PD) .....	6, 7
<b>INDAGINE CONOSCITIVA SULLE TEMATICHE RELATIVE ALL'IMPATTO DELLA TECNOLOGIA FINANZIARIA SUL SETTORE FINANZIARIO, CREDITIZIO E ASSICURATIVO</b>		Giudici Paolo, <i>Ordinario di statistica del dipartimento di scienze economiche e aziendali presso l'Università degli studi di Pavia</i> ....	3, 7, 9
		Paglia Giovanni (SI-SEL-POS) .....	6
		Pelillo Michele (PD) .....	7
		Sibilia Carlo (M5S) .....	7, 9
<b>Audizione del professor Paolo Giudici:</b>		<i>ALLEGATO: Documentazione depositata dal professor Giudici</i> .....	10
Bernardo Maurizio, <i>Presidente</i> .....	3, 6, 7, 9		

**N. B. Sigle dei gruppi parlamentari: Partito Democratico: PD; Movimento 5 Stelle: M5S; Forza Italia - Il Popolo della Libertà - Berlusconi Presidente: (FI-PdL); Articolo 1 - Movimento Democratico e Progressista: MDP; Alternativa Popolare-Centristi per l'Europa-NCD: AP-CpE-NCD; Lega Nord e Autonomie - Lega dei Popoli - Noi con Salvini: (LNA); Sinistra Italiana-Sinistra Ecologia Libertà-Possibile: SI-SEL-POS; Scelta Civica-ALA per la Costituente Liberale e Popolare-MAIE: SC-ALA CLP-MAIE; Democrazia Solidale-Centro Democratico: (DeS-CD); Fratelli d'Italia-Alleanza Nazionale: (FdI-AN); Misto: Misto; Misto-Civici e Innovatori PER l'Italia: Misto-CIpI; Misto-Direzione Italia: Misto-DI; Misto-Minoranze Linguistiche: Misto-Min.Ling.; Misto-UDC-IDEA: Misto-UDC-IDEA; Misto-Alternativa Libera-Tutti Insieme per l'Italia: Misto-AL-TIpI; Misto-FARE !-PRI-Liberali: Misto-FARE !PRIL; Misto-Partito Socialista Italiano (PSI) - Liberali per l'Italia (PLI) - Indipendenti: Misto-PSI-PLI-I.**

PAGINA BIANCA

PRESIDENZA DEL PRESIDENTE  
MAURIZIO BERNARDO

**La seduta comincia alle 14.20.**

**Sulla pubblicità dei lavori.**

PRESIDENTE. Avverto che la pubblicità dei lavori della seduta odierna sarà assicurata anche attraverso la trasmissione televisiva in differita sul canale satellitare della Camera dei deputati e la trasmissione diretta sulla *web-tv* della Camera dei deputati.

**Audizione del professor Paolo Giudici.**

PRESIDENTE. L'ordine del giorno reca, nell'ambito dell'indagine conoscitiva sulle tematiche relative all'impatto della tecnologia finanziaria sul settore finanziario, creditizio e assicurativo, l'audizione del professor Paolo Giudici, ordinario di statistica del dipartimento di scienze economiche e aziendali presso l'Università degli studi di Pavia, a cui do subito la parola.

PAOLO GIUDICI, *Ordinario di statistica del dipartimento di scienze economiche e aziendali presso l'Università degli studi di Pavia*. Grazie, presidente. Ho accolto con piacere l'invito a partecipare a questa audizione e ho preparato alcune *slide*. Cercherò di essere piuttosto sintetico nella mia relazione, per poi eventualmente approfondire alcuni punti.

Innanzitutto vorrei dirvi di che cosa mi occupo e di cosa si occupa il laboratorio *Data Science* dell'Università di Pavia, che coordino. Essenzialmente ci occupiamo di applicazioni della *Data Science*, quindi di

quelle metodologie di analisi dei *data* e dei *big data* applicate al mondo del FinTech.

Il FinTech cresce soprattutto grazie alle tecnologie di analisi dei dati e, quindi, il nostro contributo è soprattutto di tipo metodologico, ma orientato al miglioramento delle soluzioni, in particolare per quanto riguarda la protezione dei risparmiatori. Il tema finale, che vi anticipo, è che con una buona analisi dei dati è possibile tutelare i risparmiatori dagli inevitabili, maggiori rischi insiti negli strumenti del FinTech che, essendo digitali, molto flessibili e interconnessi tra loro, contengono dei rischi non solo informatici, ma anche informativi, piuttosto importanti.

La sintesi di ciò che sto per dire è che non c'è nulla di cui spaventarsi rispetto al FinTech; anzi il FinTech può essere un valore, però va governato in termini di metodologia e di strumenti.

Mi sono permesso di evidenziare alcuni, potenziali suggerimenti normativi per rendere concrete queste proposte.

Inizialmente vi riporto la definizione di *Data Science*, ossia scienza dei dati, anche se la traduzione in italiano non rende così bene come in inglese. Questa è una definizione un po' accademica, che però vorrei sottolineare, perché è importante: il *data scientist*, cioè lo scienziato dei dati, è una persona, che ancora non stiamo formando in misura adeguata nel nostro Paese, che ha tre competenze integrate: informatica, statistico-matematica e economico-finanziaria.

Abbiamo poche Università che formano *data scientist* e ancora meno che formano *financial data scientist*: una citazione che ripeto spesso è che la scienza dei dati non esiste di per sé, ma è definita in relazione all'ambito applicativo in cui viene applicata, quindi in ambito medico è una cosa,

in ambito finanziario un'altra, e non possiamo pensare che basti prendere un bravo ingegnere e fargli fare l'analisi dei dati finanziari. Occorre infatti che la persona abbia tutte le competenze che vi ho indicato.

Questo è molto importante, perché non è facile formare queste figure, che sono le più ricercate attualmente nel mercato. È difficile formarle, ma credo che nel nostro Paese abbiamo le competenze e le capacità per farlo, avendo già degli esempi concreti. Ma di questo parlerò alla fine del mio intervento.

La scienza dei dati si sviluppa, come è insito nella definizione, dal filone della statistica che, come sapete, va intesa come l'attività dell'uomo politico: lo *statistik* in tedesco era l'uomo di Stato, la persona che prendeva le decisioni sulla base di dati.

Questo è ciò che succede anche nella *data science*, nella quale, al posto dei dati, ci sono i *big data*. Qual è la differenza? Non è una questione di quantità, di volumi o di varietà dei dati, come spesso viene affermato, ma è un tema di produzione dei dati: i dati classici sono generati da noi, i *big data*, invece, sono generati dalle macchine, per esempio dai nostri telefoni quando andiamo sui social *network*, dai sensori intelligenti nelle città, dalle piccole apparecchiature elettroniche che troveremo nelle nostre case, con l'internet delle cose. Questa è la differenza.

A Pavia dal 2001 abbiamo un laboratorio che si occupava prima di *data mining* e ora di *data science*, applicato al mondo economico-finanziario. Abbiamo un gruppo con alcuni dottorandi e collaboratori che, assieme a diverse imprese e istituzioni, affronta i problemi che si presentano.

Passiamo ora al FinTech, il tema di nostro interesse. Il FinTech è, a sua volta, un'attività economica che integra, essenzialmente, la fornitura di servizi finanziari con l'utilizzo di tecnologie innovative (questa è la definizione). Queste tecnologie innovative sono normalmente *on line*, cioè vi accediamo tramite *smartphone*, e si basano sulla *data science*, sull'analisi dei dati o di *big data*.

Come è noto, le imprese FinTech sono diventate competitive, sia quelle piccole sia quelle più grandi, cioè i *big tech*, tuttavia questa maggiore competitività porta anche maggiori rischi. In particolare, due categorie di rischi molto importanti sono i *cyber risk*, cioè i rischi legati alla sicurezza informatica, e i meno noti rischi informativi, detti *scoring risk*.

Che cosa sono? Le piattaforme *peer to peer* di norma forniscono informazioni tramite una *app*, magari inserendovi il *rating* di una certa società; l'utente si deve fidare di quel *rating*, nessuno gli può confermare se è corretto o sbagliato, quindi c'è il rischio che l'utente, fuorviato dalla tipologia di *rating*, che la FinTech può inflazionare, compia delle scelte errate in termini di prestito (pensiamo al *lending*), ma anche in termini di scelte di investimento (pensiamo alla MiFID e a tutti i temi della valutazione dell'adeguatezza e della comprensione dei rischi degli utenti e, nel caso specifico, dei risparmiatori).

Inoltre ci sono rischi sistemici, perché le imprese FinTech sono, per loro natura, collegate in rete, quindi c'è una potenzialità di maggiore rischio sistemico. Ciò non significa che io ritenga che il FinTech sia negativo, perché certamente le imprese di questo settore apportano numerosi vantaggi, ma i rischi ad esse connessi vanno valutati adeguatamente.

Ho riportato un esempio concreto emerso da uno studio che abbiamo svolto recentemente, il quale evidenzia come le FinTech possono comportare rischi maggiori, ma tali rischi possono essere più facilmente valutati rispetto a quelli connessi all'attività classica delle banche, perché le imprese FinTech vivono in un mondo di dati relazionali, per esempio i dati dei *social network*, o i dati transazionali. Se riusciamo a valorizzare questi dati, allora i rischi connessi al Fintech possono essere compresi e, quindi, mitigati.

Questo è il tema del caso di studio che vi illustro, un caso concreto che mostra come dovrebbe lavorare un *data scientist* o un *team* di *data scientist*, e parla di *credit scoring* per il P2P *lending*.

Questo è un lavoro che stiamo portando avanti in collaborazione con diverse imprese FinTech, in particolare con modefinance e Borsa del credito, che sono operatori italiani. Da loro abbiamo avuto i dati in via confidenziale, abbiamo preso quelli che riguardavano il merito creditizio di un certo numero di piccole e medie imprese italiane e sulla base di questi dati abbiamo costruito un modello di relazione tra quelle imprese. Abbiamo creato relazioni che non sono le classiche reti, distretti o filiere, ma sono la loro versione moderna, basata sulle relazioni transazionali o sociali tra le imprese.

Utilizzando questi dati siamo pervenuti ai risultati successivi: in sintesi, il grafico che vi mostro evidenzia come, tra queste 727 imprese, ci siano quelle sane (in verde) e quelle andate in *default* (in rosso), come queste imprese siano molto interrelate e in queste interrelazioni giochino un ruolo determinante le imprese « cattive », le quali sono centrali.

Questo è ciò che è accaduto anche nel sistema produttivo del nostro Paese nel periodo della crisi, cioè le imprese critiche erano centrali nella loro catena di fornitura, nella loro catena di clienti. Nel grafico non ci sono nomi ovviamente, ma sappiamo chi sta dietro a ogni nodo, quindi possiamo sapere qual è l'impresa con il circolo più grande al centro, che è quella che va monitorata con più attenzione.

Abbiamo elaborato tutto ciò sulla base dei dati di bilancio relativi al fatturato di queste imprese, pertanto le relazioni che emergono sono legate al ciclo economico. Quando il ciclo va male, le imprese sono, normalmente, più correlate tra loro, quindi abbiamo un rischio sistematico piuttosto importante.

Abbiamo fatto lo stesso lavoro utilizzando un'altra variabile, che però è di tipo idiosincratico, cioè variabile: il tasso di solvibilità dell'impresa. Questo dato è legato alla singola impresa, infatti le imprese « cattive » non hanno un ruolo sistemico così importante.

Quelli che vi ho portato sono solo esempi, ma il punto importante è che essi sono esempi non solo accademici, che oltre a

poter essere pubblicati su buone riviste scientifiche internazionali possono servire anche all'utente. In particolare, un buon modello di *rating* si giudica dalla capacità predittiva, cioè da quanto è in grado di prevedere i fallimenti, prima e non dopo che essi si verificano. Spesso si fanno ragionamenti sulla crisi e su quello che è successo, *ex post*, perché è molto più difficile farlo *ex ante*.

Nelle *slide* che vi ho consegnato abbiamo applicato modelli predittivi del *default* standard, che riguardano le FinTech, ma anche le banche classiche, e modelli predittivi rafforzati con lo schema di *network* che vi ho prima descritto. La capacità predittiva passa dal 71 per cento (vuol dire che si prevede il fallimento 7 volte su 10) all'82 per cento. È un incremento considerevole.

Naturalmente è difficile arrivare al 100 per cento di capacità predittiva, perché la perfezione statistica non esiste, però c'è un effettivo miglioramento, che ci fa comprendere come possiamo proteggere il risparmiatore che investe nelle imprese valutate da un'impresa FinTech, piuttosto che quello che investe nei titoli o nel portafoglio proposti da una FinTech che fa *robo-advisory*. Possiamo proteggerlo dando una valutazione dell'algoritmo che la FinTech usa e potenzialmente migliorarlo.

Questo esempio è paradigmatico: stiamo facendo, insieme ad altre Università alle quali siamo collegati, anche altri studi in altri ambiti FinTech, ad esempio sul *robo-advisory* piuttosto che sul mondo delle criptovalute, che pone diversi temi, sui quali adesso non mi soffermo, ma nei quali anche il metodo statistico può essere molto importante.

Passando alla penultima *slide* che vi illustrerò, essa indica alcuni suggerimenti che mi sono permesso di sviluppare sulla base della mia esperienza personale e del gruppo di persone con cui lavoro, anche a livello internazionale. Il suggerimento è, essenzialmente, che un buon metodo di analisi dei dati, chiamiamolo *Data Science*, può migliorare la protezione dei risparmiatori concretamente e non solo teoricamente.

Per fare questo – nulla è a costo zero – è necessario un investimento in capitale umano, cioè la formazione di *data scientist* adeguati, non degli informatici o degli statistici o degli esperti in finanza, ma persone che abbiano tutte e tre queste competenze.

Ciò andrebbe realizzato in modo concreto, non teorico, quindi i suggerimenti che vi do, anche sulla base di situazioni passate o di altri Paesi, potrebbero essere: 1) esternalizzare le attività dei laboratori di *data science* presso *sandbox* o *innovation hub* che vogliamo eventualmente costituire, perché sarebbe molto importante mettere in contatto diretto l'università con le imprese e con il regolatore, favorendo una ricerca concreta e meno astratta; 2) indirizzare finanziamenti che già esistono su un tema specifico di *data science*: è stato già fatto per *big data* rispetto ai corsi di dottorato; 3) per fare una buona analisi dei dati – credo se ne sia già parlato in altre audizioni svolte nell'ambito dell'indagine conoscitiva – è importante che la materia prima, cioè i dati, sia di buona qualità.

Il mondo del FinTech può avere dati di tipo transazionale, non solo, quindi, i dati della relazione fra le banche e le imprese, sul modello della Centrale dei rischi, ma dati ulteriori, che riguardano le transazioni tra le imprese. Alcune società stanno costruendo questo e vengo ora da un convegno presso l'ABI (Associazione bancaria italiana) dove si è parlato anche di questo. Sarebbe molto importante, a mio avviso, che il Paese incentivasse lo sviluppo di un *database* pubblico di transazioni, prima che l'avvento della Direttiva PSD2, relativa ai servizi di pagamento nel mercato interno, e del Regolamento UE in materia di protezione dei dati (GDPR) ci tolgano questa possibilità e magari la indirizzino verso società *big tech* che probabilmente non faranno solo gli interessi del nostro Paese.

Grazie per l'attenzione.

PRESIDENTE. Grazie, professore. Do la parola ai colleghi che desiderino intervenire per porre quesiti o formulare osservazioni.

SEBASTIANO BARBANTI. Grazie, professore, per l'audizione e le informazioni

che ci ha dato, che si collegano a una delle nostre ultime audizioni svolte nell'ambito dell'indagine conoscitiva, nel corso della quale si è parlato di rischio di credito e di *rating*, analizzando come viene costruito nel FinTech.

Con il suo intervento abbiamo fatto un passo avanti in questa direzione, scendendo sulla verticale di come il *rating* può essere creato e applicando le metodologie sottostanti; abbiamo quindi toccato un argomento molto interessante non solo per noi, ma per qualunque *regulator*, cioè il tema della stima puntuale del rischio di credito, ma anche della correlazione tra i vari rischi di credito delle aziende. Se pensiamo alle gravi conseguenze della crisi successiva al caso Lehman Brothers, è evidente che c'è stato un rischio sistemico che si è diffuso. Penso quindi che la stima di questo rischio sia fondamentale.

Mi sembra che oggi abbiamo l'opportunità di compiere questa stima con le metodologie che ci ha illustrato e che, come giustamente ha detto, rischiano di scontrarsi con la normativa, sia quella recata dal regolamento dell'Unione europea GDPR, recante misure in materia di protezione dei dati, sia la direttiva PSD2, relativa ai servizi di pagamento nel mercato interno. Le chiedo quindi: come possono correlarsi queste due esigenze?

Al punto c) dei suoi suggerimenti giustamente lei ci chiede di incoraggiare la costruzione di un *network database* che nel Regno Unito si è riusciti a fare riunendo insieme i dati posseduti da alcune banche, in maniera anonima, consentendo la fruibilità di alcuni dati, interni, per l'addestramento di reti neurali piuttosto che per la creazione di modelli statistici sottostanti.

È molto interessante e condivido in pieno il tema della collaborazione, soprattutto in questo momento, quindi le chiedo come sia possibile realizzarla all'interno di un *hub* e se ci possano essere delle proposte normative in grado di semplificarne l'*iter*, dalla costruzione dell'*hub* all'adesione ad esso, fino al lavoro da svolgere, praticamente, all'interno dell'*hub*. Grazie.

GIOVANNI PAGLIA. La ringrazio anch'io per la sua relazione. Rispetto alla



raccolta dei dati (parliamo di merito creditizio e di sistemi per la valutazione del rischio in relazione alle imprese) quali sono le fonti che si utilizzano o dovrebbero essere utilizzate, a suo avviso, per la raccolta? Suppongo che non saranno solo i dati evidenti, cioè i dati di bilancio, ma che si utilizzino anche altre fonti.

MICHELE PELILLO. Due domande, professore. La prima è una mia curiosità: qual è il nome del corso di studi in laurea magistrale?

Mi ha colpito molto il discorso sulla capacità predittiva. Questi modelli di *network* da voi proposti sono già stati adottati da qualche fondo importante?

CARLO SIBILIA. La mia domanda somiglia a quella appena posta dal collega Pelillo. Queste operazioni predittive in senso stretto, secondo lei, possono essere implementate, anche a livello di vigilanza bancaria vera e propria? Creando un *network* delle transazioni, cioè l'archivio di cui lei giustamente parlava, forse si potrebbe, con un progetto a lungo termine, creare un sistema che possa anche autovigilarsi.

Le stesse banche o gli organismi di vigilanza, come la Banca d'Italia, a volte danno informazioni non corrette e non coerenti con le reali situazioni patrimoniali, le quali potrebbero trovare riscontro in transazioni fatte con le imprese. In questo modo si avrebbe, anche se non in tempo reale, un dato statistico talmente consistente da riuscire a metterci al riparo da grandi disastri, sia dal punto di vista delle banche (su questo siamo tutti d'accordo, a prescindere dalle idee politiche) sia delle imprese.

Ricordo i crac Parmalat e Cirio: se avessimo disposto di questa tecnologia, con un meccanismo del genere avremmo potuto salvaguardare quelle situazioni in modo diverso. Mi chiedo quindi se ci siano, in questo momento, delle collaborazioni tra chi usa questi metodi e gli organismi di vigilanza come, ad esempio, la Banca d'Italia e la Consob.

Secondo lei, infine, quali devono essere gli *step* da portare avanti, a livello legisla-

tivo, per creare un archivio di transazioni, connesso a quella fase di sperimentazione e creazione dell'*hub* di cui lei ci ha parlato?

SEBASTIANO BARBANTI. Mi perdoni, ho un'ultima domanda: ho notato che avete utilizzato i dati insieme a *modefinance* e alla Borsa del credito e mi chiedevo se, tra i dati che avete utilizzato per questo miglior *fitting*, cioè per il miglioramento della capacità predittiva, ci fossero anche i dati della Centrale rischi. Nel corso dell'audizione svolta nell'ambito dell'indagine conoscitiva, infatti, *modefinance* ci ha fatto presente che, nonostante sia l'unica agenzia accreditata dall'ESMA (*European Securities and Markets Authority*) in Italia, non ha la possibilità di accedere ai dati della Centrale rischi, che considero fondamentali. Questo significherebbe incrementare la capacità predittiva, disponendo di un *set* di dati quanto più ampio possibile. Grazie.

PRESIDENTE. Lascio la parola al professor Giudici per le conclusioni.

PAOLO GIUDICI, *Ordinario di statistica del dipartimento di scienze economiche e aziendali presso l'Università degli studi di Pavia*. Ringrazio tutti i membri della Commissione per le domande molto precise e puntuali, e spero di essere sintetico nelle risposte, anche se non è facile, perché si tratta di molte domande, tutte davvero importanti.

Parto dall'ultima domanda posta dall'onorevole Barbanti. Nell'analisi che lei citava non abbiamo utilizzato i dati della Centrale dei rischi perché, appunto, tali dati non sono pubblicamente disponibili. Se lo fossero, potrebbero forse – ma questo ovviamente è da valutare concretamente – migliorare ulteriormente la capacità predittiva. Va da sé che, quanti più dati disponiamo, e se utilizziamo buoni modelli, tanto più questa capacità può migliorare.

Quanto alla correlazione tra la stima del rischio e la *privacy*, nell'ambito di quanto previsto dal citato Regolamento GDPR, questo è un grande tema. In altri contesti, per esempio in medicina, si tratta di una que-

stione nota: è noto il fatto che i dati debbano essere prodotti in forma aggregata, quindi anche i modelli di *network* non devono andare sul singolo caso, bensì produrre dati aggregati, che nascondono il dato che è dietro.

Dico di più: reputo sia il Regolamento GDPR sia la Direttiva PSD2 come un'occasione per acquisire ulteriori informazioni da altri *player*, anche internazionali. Ciò anche perché c'è la questione del diritto all'oblio, che è un tema estremamente importante anche nel citato Regolamento sulla protezione dei dati. Ciò significa che le analisi storiche prodotte dalle banche potrebbero ridurre la loro capacità predittiva perché in qualunque momento il dato storico potrebbe essere eliminato. L'importanza del dato transazionale e del *network* emerge quindi ancora di più. Dunque le due questioni si sposano, sono strettamente collegate.

Sul tema della modalità collaborativa, credo che le importanti iniziative legislative recentemente adottate, come quella relativa al dottorato industriale e al credito d'imposta per le imprese che investono in attività di ricerca e sviluppo presso università, siano due pilastri fondamentali sui quali poggiare l'idea di collaborazione, in termini di *sandbox* o di *FinTech hub*. In particolare credo che il nostro dottorato industriale, i cui corsi vengono realizzati, per un piccolo numero di studenti, tra Pavia, Bergamo e Milano, funzioni bene, perché il dottorando passa metà del suo tempo all'università e metà presso la *FinTech* o la banca con la quale si collabora, e in questo modo riesce a imparare concretamente.

Credo che questi possano essere strumenti validi.

Le fonti del *credit scoring* sono ovviamente numerose, e naturalmente è tanto meglio quanto più esse sono di qualità, perché altrimenti «spazzatura dentro — spazzatura fuori», non possiamo pensare che il modello risolva il problema della qualità dei dati. Ci sono varie fonti, i dati di bilancio hanno il loro limite connaturato al fatto che il bilancio non viene redatto con quella finalità, mentre ci sono dati impor-

tanti che provengono dalle transazioni che il cliente compie con la banca. Ci sono inoltre i dati che le banche usano abitualmente per lo *scoring*, cioè i dati del cosiddetto *self-assessment*: il responsabile di filiale valuta il cliente in modo qualitativo, sulla base della propria esperienza.

Questa attività verrà sviluppata ancora di più in futuro: torniamo cioè al passato, e l'importanza della relazione, fatta tramite un *social network*, torna ad essere importante, così come lo era nelle filiali di un tempo. La banca territoriale è quindi ancora un valore importante, anzi vorrei trasmettere questo messaggio: non sono solo le banche grandi che possono avere un ruolo nell'ambito dell'ecosistema del *FinTech*; anzi, le banche più piccole, territoriali, possono percorrere il loro cammino in questo nuovo mondo, magari anche meglio di altre.

Altre fonti sono la Centrale dei rischi, come si è detto, che è importante ma corrisponde a un modo di pensare legato al dato storico. Servono quindi altre Centrali dei rischi, come quella delle transazioni di cui si parlava prima.

Tutte queste fonti di dati vanno integrate in un unico modello di analisi, e ciò si può fare valutando se quel nuovo *database* è utile in termini di capacità predittive, perché se non aggiunge nulla o peggiora la capacità predittiva, lo si può anche eliminare.

Per quanto riguarda il nome del percorso magistrale è una bella domanda perché, come ho detto all'inizio, per me la *data science* non esiste di per sé, quindi non farei corsi di *data science*, che non vuol dire nulla: essa va applicata a qualcosa. Noi abbiamo ad esempio un corso di laurea magistrale in lingua inglese, *Economics, Finance and international integration*, dove *data science* non è nel titolo, ma è uno degli insegnamenti: uno strumento, che deve essere utilizzato in economia e finanza. Credo infatti che il punto di partenza debba essere l'aspetto applicativo.

Per quanto riguarda la domanda relativa all'utilizzo delle tecnologie di *network* da parte di altre istituzioni, la risposta è «assolutamente sì». Domani parteciperò a



un convegno a Francoforte, il primo convegno di modelli di *network* in finanza, dove ci saranno più di 100 operatori bancari, finanziari e tecnologici e si discuterà proprio di questi modelli, che sono stati implementati fino a oggi da istituzioni importanti, come la Banca dei regolamenti internazionali e diverse banche centrali, proprio per valutare il rischio sistemico.

Riprendendo quanto già affermato, il rischio sistemico si lega naturalmente al rischio di credito del singolo imprenditore, e questo è un fatto molto importante. Noi abbiamo partecipato a questi sviluppi e io stesso l'anno scorso ho lavorato presso la Banca dei regolamenti internazionali, in qualità di *research fellow*, su questo tipo di modelli.

Alla domanda se gli algoritmi di *network* siano implementati a livello di vigilanza, credo di aver già parzialmente risposto, la risposta è sì, presumibilmente ciò verrà fatto ancora di più, ci sono anche degli operatori tecnologici, i cosiddetti *data provider*, che nel nostro Paese si stanno muovendo in questo ambito, costruendo quelle che vengono chiamate le «reti di filiere innovative».

Per quanto riguarda invece gli strumenti legislativi, in particolare con riferimento al *database*, la FCA (*Financial conduct authority*) inglese ha fatto questo, cioè, tramite un *sandbox* di imprese, ha costruito un *database* di imprese collegate alle Fin-Tech ed è pronta a compiere delle analisi tramite questo *database*.

Mi sono anche permesso di chiedere loro, durante un convegno, se fosse possi-

bile analizzare il *database* e mi hanno risposto di sì, salvo poi precisarmi che si trattava di dati riservati e avevano già stretto un accordo con l'Università di Cambridge per svolgere quello stesso lavoro. Però, magari a livello nazionale, potremmo realizzare qualcosa di analogo.

Spero di aver risposto a tutte le domande.

CARLO SIBILIA. Le tempistiche relative ai *sandbox* ?

PAOLO GIUDICI, *Ordinario di statistica del dipartimento di scienze economiche e aziendali presso l'Università degli studi di Pavia*. Credo un paio d'anni, non di più, complessivamente. Si tratta di un tempo piuttosto breve. Poi c'è l'esempio di Singapore e dell'Australia, che sono stati anche più veloci, però credo che quello sia il tempo impiegato mediamente.

PRESIDENTE. Ringrazio il professor Giudici, la cui audizione odierna ha suscitato molto interesse. Autorizzo la pubblicazione in allegato al resoconto stenografico della seduta odierna della documentazione consegnata dal professore (*vedi allegato*) e dichiaro conclusa l'audizione.

**La seduta termina alle 14.55.**

---

*Licenziato per la stampa  
il 17 gennaio 2018*

---

Data Science  
FinTech  
Caso studio  
Suggerimenti  
Appendice

# Data science e FinTech

## Audizione in Commissione Finanze

### Camera dei Deputati, Roma, 25 ottobre 2017

Paolo Giudici

Professore Ordinario di **Data Science** e Statistica  
Dipartimento di Scienze Economiche ed Aziendali, Università di Pavia  
giudici@unipv.it; Twitter: @stateconomist; Facebook: Data Science Network



Giudici

Data science e FinTech Audizione in Commissione Finanze

ALLEGATO

## Premessa

Una nota simile alla presente e' stata sottoposta a:  
*Fintech EU: European Commission public consultation on FinTech, Giugno 2017.*  
E' stata inoltre discussa nel corso del ciclo di incontri organizzati dalla CONSOB nell'anno in corso, in collaborazione con le principali Fintech e Università Italiane.

Nei prossimi giorni, verrà inviata come contributo alla discussione dei seguenti documenti:

- *BIS Consultative Document on sound practices on the implications of fintech developments for banks and bank supervisors;*
- *EBA Discussion Paper on financial technology consultation;*
- *ECB Guide to the assessment of fintech credit institution license applications.*



Data Science  
FinTech  
Caso studio  
Suggerimenti  
Appendice

## Data Science

Attività che integra tre componenti:

- La selezione e l'elaborazione di grandi masse di dati (Informatica)
- L'individuazione di regole empiriche, basate sui dati: regolarità e relazioni (Statistica)
- L'interpretazione di regole empiriche, rispetto al contesto applicativo di interesse (Economia e Finanza)

"Data science does not exist; Data ScienceS do" (Giudici, 2017)



Giudici

Data science e FinTech Audizione in Commissione Finanze

Data Science  
FinTech  
Caso studio  
Suggerimenti  
Appendice

## Data Science - origine

- 1700s: statistica (statistik) = statistica + politica
- 1800s: inferenza statistica = statistica + scienze
- 1980s: statistica + informatica
- 2000s: **data mining** = informatica + statistica + contesto applicativo su DATA (generati "consapevolmente")
- 2010s: **data science** = informatica + statistica + contesto applicativo su BIG DATA (generati in automatico)



Giudici

Data science e FinTech Audizione in Commissione Finanze



Data Science  
FinTech  
Caso studio  
Suggerimenti  
Appendice

## Data Science - il nostro contributo

- Feb 2001: Istituzione del laboratorio di DATA MINING, presso la Facoltà di Economia @unipv.it
- Feb 2016: Il laboratorio di data mining si trasforma in laboratorio di DATA SCIENCE, presso il Dipartimento di Economia @unipv.it, coordina un network di discussione e ricerca internazionale e istituisce un corso di studi dedicato, sia a livello di Laurea Magistrale che di Dottorato (entrambi in lingua inglese).



Giudici

Data science e FinTech Audizione in Commissione Finanze





- Il termine FinTech (FINancial TEChnology) descrive un'attività economica che **integra** la fornitura di servizi finanziari con l'utilizzo di tecnologie innovative (tipicamente on-line, e basate su processi di **data science**)
- Gli sviluppi della data science hanno consentito alle FinTech di fornire servizi **competitivi** rispetto ai servizi "tradizionali", nei vari ambiti: creditech, paytech, wealhtech, insurtech,...
- Per loro natura, tuttavia, le FinTech possono implicare **maggiori rischi** per gli utenti, quali **cyber risk**: rischio di perdite finanziarie dovute a "fallimenti" dei sistemi IT e **scoring risk**: rischio di perdite finanziarie dovute alla fornitura di erronee informazioni (es. rating). Entrambi possono essere amplificate dai **rischi sistemici**, causati dall'ambiente interconnesso delle FinTech.



## Fintech - il nostro contributo

- Il laboratorio di Data Science è composto da 3 Professori di Ruolo, 8 studenti di dottorato, 50 studenti Master, e da una rete internazionale di circa 3,800 partecipanti. Effettua attività di ricerca, di base ed applicata, ed attività di formazione sul campo, in collaborazione con diverse Istituzioni Finanziarie, FinTech e Regolatori, a livello nazionale ed internazionale.
- L'attuale obiettivo del laboratorio è lo sviluppo e l'applicazione di metodi di data science per la valutazione e la mitigazione dei rischi delle FinTech Ciò, in particolare, valorizzando le informazioni provenienti dalle diverse reti relazionali che generano "big data": reti sociali, reti transazionali.
- Esempi di lavori in corso: modelli di credit scoring per il P2P lending; modelli di valutazione dell'adeguatezza dei profili di



## Credit scoring per il P2P lending

- Abbiamo "raccolto" dati, sulla base di collaborazioni di ricerca con: Lending Club, Mode Finance, Borsa del Credito, N26 (in corso).
- Abbiamo applicato ai diversi dati disponibili differenti metodi di apprendimento di regole statistiche, ed abbiamo scelto quelli maggiormente predittivi del rischio di credito (default).
- Nel seguito, ci riferiamo ai dati pervenuti da Mode Finance, riguardanti 727 piccole e medie imprese che hanno richiesto un finanziamento tramite una piattaforma P2P Italiana. La percentuale di default nel campione osservato è pari al 23%.



## Risultati: rete "statistica" di imprese - sistematica

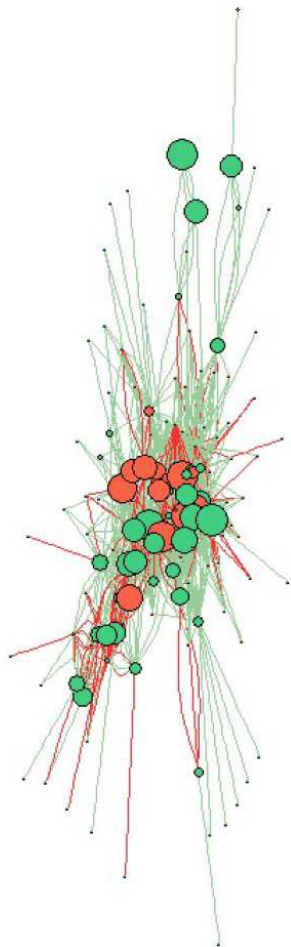


Figure: La rete statistica di cui sopra è stata stimata sulla base della correlazione fra i fatturati relativi delle imprese, nel periodo considerato (post-crisi). Ogni nodo corrisponde ad un'impresa; il colore verde indica un'impresa "sana" ed il rosso una in "default". La dimensione dei nodi è funzione della centralità della impresa considerata nel network complessivo.





## Risultati: rete "statistica" di imprese - idiosincratica

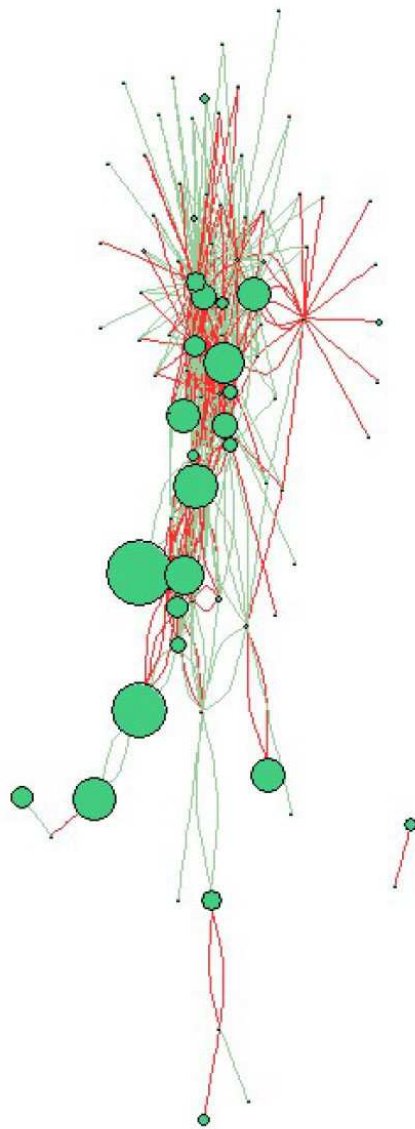


Figure: La rete statistica di cui sopra è stata stimata sulla base della correlazione fra i tassi di solvibilità delle imprese, nel periodo considerato (post-crisi). Ogni nodo corrisponde ad un'impresa; il colore verde indica un'impresa "sana" ed il rosso una in "default". La dimensione dei nodi è funzione della centralità della impresa considerata nel network



## Capacità predittiva: modelli standard

Variable	Estimate	P-value	Significance
Intercept	-3.39	0.011	*
Solvency ratio	0.01	0.539	
Debt to equity ratio	-0.07	0.517	
Current ratio	0.21	0.032	*
Cash over total assets	-2.51	0.579	
Return on equity	-0.08	0.008	**
Return on assets	0.01	0.963	
Return on Capital Employed	0.09	0.044	*
Coverage	-0.01	0.875	
Activity ratio	-1.92	0.001	***
Capacità predittiva (AUROC)			<b>0.71</b>



Table: The estimated baseline regression model



## Capacità predittiva: modelli network da noi proposti

Variable	Estimate	P-value	Significance
Intercept	-1.53	0.033	*
Solvency ratio	-0.02	0.012	*
Debt to equity ratio	-0.00	0.576	
Current ratio	0.24	0.072	*
Cash over total assets	1.08	0.443	
Return on equity	-0.11	0.000	***
Return on assets	0.02	0.876	
Return on capital employed	0.01	0.212	
Coverage	0.02	0.248	
Activity ratio			
Degree Centrality	0.01	0.026	*
Closeness	1.05	0.002	**
Capacità predittiva (AUROC)			<b>0.82</b>



## Interpretazione economica

- I modelli di network proposti migliorano la capacità predittiva dei default di circa 11 punti percentuali. In aggiunta, forniscono utili informazioni su quali imprese siano da monitorare attentamente, in termini idiosincratici e di potenziale contagio. In generale, i modelli proposti proteggono meglio gli utenti dei servizi P2P, rispetto agli standard attualmente in uso.
- Ci attendiamo che ulteriori esperimenti e/o ulteriori dati (in particolare, da reti transazionali) possano migliorare ulteriormente la performance dei modelli proposti.
- Risultati simili emergono dalle attività di analisi in corso su altri temi FinTech: roboadvisory, cryptocurrencies, cyber risks.



## Suggerimenti

- I metodi di Data Science possono contribuire alla protezione degli utenti (risparmiatori) delle FinTech, **riducendo i rischi e preservando la stabilità finanziaria**.
- Lo sviluppo dei metodi di data science richiede **investimenti** in formazione e ricerca, specialmente in modalità collaborativa fra le Università, l'Industria Finanziaria ed i Regolatori.
- Sugeriamo i seguenti investimenti "leggeri": a) **esternalizzare parte delle attività dei laboratori accademici di data science** presso i distretti, innovation hub e sandbox che verranno eventualmente costituiti in tema di FinTech; b) **bandire finanziamenti di ricerca dedicati** che prevedano obbligatoriamente la modalità collaborativa di cui sopra; c) **incoraggiare la costruzione di un "network database"**, nazionale o Europeo, contenente dati



Data Science  
FinTech  
Caso studio  
Suggerimenti  
Appendice

## Alcune recenti pubblicazioni del laboratorio di data science

- (2017) Pejman Abedifar, Paolo Giudici, Shatha Hashem. Heterogeneous market structure and systemic risk: evidence from dual banking systems. *Journal of Financial Stability*.
- (2017) Paolo Giudici, Peter Sarlin, Alessandro Spelta. The interconnected nature of financial systems: direct and common exposures. *Journal of Banking and Finance*.
- (2017) Paolo Giudici, Laura Parisi. Sovereign risk in the Euro area: a multivariate stochastic process approach. *Quantitative Finance*.
- (2016) Paolo Giudici, Alessandro Spelta. Graphical network models for international financial flows. *Journal of Business and Economic Statistics*, 34 (1), pp. 126-138.
- (2017) Paolo Giudici, Laura Parisi. Correlation networks to measure the systemic implications of bank resolution. Submitted to *Review of Finance*
- (2017) Paolo Giudici, Laura Parisi. Corisk: measuring systemic risk through default probability contagion. Submitted to *Quantitative Finance*
- (2017) Stefan Avdjiev, Paolo Giudici, Alessandro Spelta. Measuring contagion risk in international banking. Bank for International Settlements research paper. Submitted to *Journal of Financial Stability*
- (2017) Paolo Giudici, Branka Hadji Misheva. Scoring models for P2P lending platforms: a network approach. Submitted to *Journal of Banking and Finance*



Giudici

Data science e FinTech Audizione in Commissione Finanze



\*17STC0027800\*