

11 Marzo 2011



UNIVERSITÀ DI PISA
Facoltà di Economia
Facoltà di Scienze, Matematiche, Fisiche e Naturali
Corso di Laurea Magistrale in Informatica per l'economia e per l'azienda
(Business Informatics)

TESI DI LAUREA
SAS E PENTAHO: UN CASO DI STUDIO NEL MONDO DEL
RICICLAGGIO DEL DENARO

RELATORE
Prof. Nicola Ciaramella

CANDIDATO
Francesco Fontana

ANNO ACCADEMICO 2009-2010

Riassunto

Il riciclaggio del denaro sporco consiste nel reinserimento fondi provenienti da fonti illegali all'interno del mercato mascherandoli con attività lecite. Durante gli anni sono state molte le tecniche adottate dai criminali al fine di mascherare l'origine di questo denaro, ma altrettanto valide sono state le contromisure adottate dalla legge per contrastarle. Con l'avvento dell'informatica sono stati sviluppati dei software specifici per l'identificazione di movimenti bancari sospetti. Questo lavoro introduce le tecniche di analisi avanzata dei dati, tipiche della Business Intelligence, in uno di questi programmi, GIANOS®©, per facilitare l'identificazione delle transazioni anomale. Lo studio esamina delle possibili implementazioni di analisi volte a segnalare con maggiore accuratezza le transazioni sospette. A questo scopo è stata realizzata un'applicazione dimostrativa degli algoritmi proposti ed è stata creata un'interfaccia verso due delle principali suite di Business Intelligence: SAS e Pentaho.

Indice

Introduzione	5
Scopo del progetto	5
1 L'importanza delle scelte	7
1.1 Il tramonto degli Oracoli	7
1.2 Dai DSS alla BI	8
2 Il problema del riciclaggio	10
2.1 L'evoluzione della normativa	10
3 Lo stato dell'arte	13
3.1 L'Archivio Unico Informatico	13
3.2 Il software GIANOS®©	15
3.3 La struttura di GIANOS®©	15
3.3.1 GIANOS®© - Know Your Customer	15
3.3.2 GIANOS®© - Gestione Profili di Rischio	16
3.3.3 GIANOS®© - Valutazione Inattesi	18
3.4 Il funzionamento di GIANOS®©	18
3.5 I comportamenti	20
4 Un po' di matematica	21
4.1 Le ciclicità nel comportamento	21
4.2 Anomalia dell'importo	22
4.3 Il test statistico	24
4.4 Anomalia del periodo	25
4.5 Pesare un inatteso GIANOS®©	25
4.6 Analisi dei risultati	26
5 Gli strumenti per l'analisi: SAS e Pentaho	28
5.1 SAS	28
5.2 Pentaho	29
5.3 L'integrazione	29

6	Lo sviluppo dell'applicazione	30
6.1	Connessione al database	30
6.2	Le entità	31
6.3	L'analisi dell'entità	31
6.4	Creazione dell'output	33
7	Conclusioni	36
8	Sviluppi futuri	37
9	Ringraziamenti	38
A	L'Archivio Unico Informatico	39
B	Gli indici di anomalia	42

Introduzione

Negli ultimi mesi ho avuto la possibilità di lavorare in vari progetti collegati al mondo dell'antiriciclaggio. I progetti spaziavano dalla registrazione delle transazioni nell'Archivio Unico allo studio della struttura e delle funzionalità di GIANOS®©, un software per l'individuazione di transazioni sospette e la loro segnalazione.

Abbiamo così individuato la possibilità di introdurre alcune tecniche informative legate alla Business Intelligence, come la predizione tramite regressione e i test statistici, all'interno del *workflow* bancario.

L'idea di fondo di questo lavoro è fornire uno strumento aggiuntivo agli operatori delegati ad analizzare le centinaia di transazioni segnalate come inattesi da GIANOS®©. Per quanto nessun progetto potrà mai sostituire completamente l'intervento umano in questioni tanto delicate, sarà però possibile facilitare il compito di analisi agli operatori grazie a un'analisi a posteriori delle transazioni segnalate.

I capitoli seguenti analizzeranno i passi che hanno portato allo sviluppo di un tool di analisi degli inattesi di GIANOS®© integrato con le suite di Business Intelligence SAS e PENTAHO.

Il testo in particolare richiamerà nei primi due capitoli alcune nozioni storiche sulla Business Intelligence e sulla normativa antiriciclaggio italiana. Nel capitolo successivo verrà esaminato il software GIANOS®© e il suo funzionamento. Il capitolo seguente evidenzierà gli aspetti matematici alla base del procedimento adottato e seguirà un capitolo introduttivo su SAS e Pentaho. Infine l'ultimo capitolo mostrerà alcuni particolari del programma.

Scopo del progetto

L'idea di fondo del progetto è stata creare uno strumento informativo non basato su anomalie funzionali, come ad esempio le strategie più comuni di riciclaggio del denaro, ma su un approccio statistico, orizzontale rispetto ai controlli attualmente presenti.

Il progetto consiste nell'andare a confrontare le anomalie segnalate da GIANOS®© con i dati presenti sul soggetto nell'Archivio Unico, cercando di individuare se il movimento bancario è anomalo temporalmente o economicamente per il soggetto che ha effettuato l'operazione (Fig. 1).

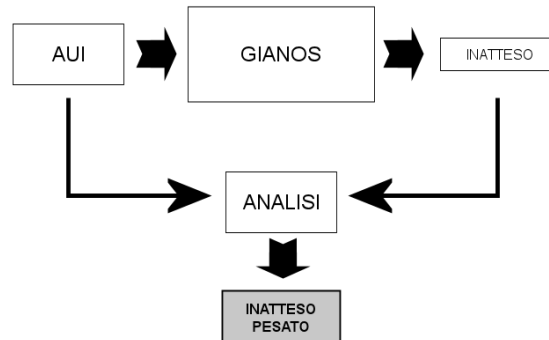


Figura 1: Struttura dell'analisi

Inoltre è stata realizzata una strategia di valutazione dei risultati in grado di fornire statistiche sul funzionamento degli algoritmi e utili informazioni sulla correlazione tra gli indicatori utilizzati.

Capitolo 1

L'importanza delle scelte

1.1 Il tramonto degli Oracoli

Creso, re dei Lidi e di altre genti, dato che ritiene che questi oracoli siano gli unici validi fra gli uomini, vi diede dei doni degni dei responsi e ora vi chiede se debba fare una spedizione contro i Persiani e se si debba fare alleato qualche esercito.

Erodoto (I, 53)

Così Creso, sovrano del regno di Lidia, cercò negli oracoli una risposta. Prima dell'avvento del libero arbitrio, l'unico modo in cui l'uomo poteva cercare di attuare delle scelte era conoscere la volontà degli dei e, come Creso insegna, le loro parole non sempre erano semplici da interpretare.

Con il passare dei secoli, l'uomo trovò strumenti sempre più efficaci per poter gestire il rischio di una nuova scelta, abbandonando gli oracoli e avvicinandosi maggiormente alle scienze. Nel 1654 Antoine Gombaud, Chevalier de Méré, si lamentò con un suo amico, Blaise Pascal, dei risultati di un gioco a dadi. Gombaud sosteneva di avere calcolato che ottenere almeno un 6 in 4 lanci equivaleva ad ottenere una coppia di 6 in 24 lanci, ma nonostante questo le sue perdite erano maggiori delle sue vincite. Così Pascal cominciò un lungo scambio epistolare con il suo amico Pierre de Fermat, delineando la concezione frequentista del calcolo delle probabilità.

Fu un secolo di grandi progressi nel campo della probabilità e della statistica: Gottfried von Leibniz, Jacob e Daniel Bernoulli, Abraham de Moivre e molti altri scienziati elaborarono e approfondirono le regole e gli strumenti tuttora utilizzati.

Quasi cento anni dopo, un ministro presbiteriano britannico, Thomas Bayes, dimostrò come fondere matematicamente le informazioni conosciute con nuove informazioni, creando una nuova corrente probabilista, la cosiddetta statistica bayesiana.

È interessante notare come quasi tutti gli strumenti utilizzati nel *risk management* e nell'analisi delle decisioni risalgano a scoperte avvenute tra il 1654 e il 1760.

Le uniche eccezioni sono la *regressione*, scoperta da Francis Galton nel 1875, e la cosiddetta *frontiera del portafogli* di Harry Markowitz, elaborata nel 1952 [Ber98].

1.2 Dai DSS alla BI

La storia del primo DSS risale alla Seconda guerra mondiale, per contrastare una delle più letali armi dell'epoca: il bombardiere. "The bomber will always get through", pronuncerà Stanley Baldwin nel 1932 davanti al Parlamento inglese. L'aereo infatti viaggiava a 320 km/h, rilasciava le bombe e scompariva all'orizzonte prima che l'aviazione fosse in grado di individuarlo, mandare in volo gli aerei e abbatterlo. Dopo l'invenzione del radar la situazione migliorò sensibilmente, ma il problema di gestire ed elaborare l'enorme quantità di informazione rimase.

Fu così che nel 1948 nacque il *Project Whirlwind*, un progetto per raccogliere, integrare ed elaborare l'enorme flusso di dati dei radar. Il sistema analizzava la traiettoria del bersaglio e comunicava i dati agli operatori. Nel 1958 cominciò quindi la sua implementazione con la costruzione dei primi SAGE (Semi-automatic Ground Environment), che non verrà terminata prima del 1963. I SAGE erano composti da due computer: ciascuno delle dimensioni di $2000m^2$, integravano 55.000 valvole e pesavano 275 tonnellate. Le varie stazioni radar comunicavano con l'impianto SAGE tramite modem e le informazioni elaborate venivano visualizzate in appositi schermi, con cui gli operatori potevano interagire tramite una *light gun* [SAG09].

Nel 1964 l'arrivo sul mercato dell'IBM SYSTEM 360 consentì alle grandi aziende l'implementazione del Management Information System (MIS), grazie al quale i manager potevano ricevere *report* periodici strutturati. Il sistema non permetteva alcun supporto interattivo per aiutare i manager nel prendere decisioni.

Nel 1971, in un articolo di Gorry e Scott Morton, venne coniato il termine *Decision Support System*, definito come un sistema di supporto informativo per decisioni semi-strutturate o non strutturate. Nello stesso anno T.P. Gerrity Jr. scrisse un articolo intitolato "The Design of Man-Machine Decision Systems: An Application to Portfolio Management" sulla possibile creazione di un sistema di supporto per aiutare gli investment managers nell'amministrazione giornaliera del portfolio dei clienti.

Grazie a John D.C. Little sempre negli anni Settanta i DSS aprirono le loro frontiere al marketing con BRANDAID, un sistema che supportava le decisioni non solo riguardanti il prodotto, ma anche la promozione, il pricing e la pubblicità. Lo stesso Little sviluppò un linguaggio di modellazione finanziario e di marketing noto come EXPRESS.

Il decennio del 1970 segnò il passaggio dei DSS da un ambito sperimentale a un ambito accademico, con lo sviluppo di teorie, libri di testo e articoli scientifici. In questo periodo vennero pubblicate vere e proprie pietre miliari, come il *DSS textbook* di Keen e Scott Morton (1978) o il *Building Effective Decision Support Systems* di Ralph Sprague e Eric Carlston (1982).

Il DSS si dimostrò uno strumento potente e duttile, applicabile in decine di ambiti. Dal più semplice *model-driven DSS*, sviluppato per aiutare i managers ad analizzare una situazione, ai *communications-driven DSS*, come i groupware e le videoconferenze. Molti diversi sistemi vennero creati nel corso degli anni, come i *data-driven DSS*, sviluppati per accedere e manipolare dati organizzati in serie temporali, o i *knowledge-driven DSS*, orientati verso problemi specialistici.

Grazie alla creazione del WWW, in tempi più recenti sono stati realizzati *web-based DSS*, in grado di condividere enormi quantità di informazioni e di convertirle in efficaci strumenti decisionali [Pow07].

Negli anni Ottanta il mondo delle imprese comprendeva appieno l'importanza della giusta informazione nel giusto momento alla giusta persona, ma i dati erano frammentati e i database di difficile accesso. La soluzione di questo problema venne da due ricercatori dell'IBM, Barry Devlin e Paul Murphy che nel 1988 realizzarono il primo Data Warehouse. L'idea alla base della ricerca era utilizzare un'alta rindondanza dei dati in modo da renderli più rapidamente accessibili.

Nel 1989 Howard Dresner propose il termine *Business Intelligence* per rappresentare l'evoluzione dei DSS in sistemi volti a migliorare le decisioni di business.

Capitolo 2

Il problema del riciclaggio

2.1 L'evoluzione della normativa

Nel nostro Paese la prima individuazione del riciclaggio come reato a parte risale al 21 marzo del 1978, quando venne introdotta una prima formulazione dell'articolo 628bis.

Fuori dai casi di concorso nel reato, chiunque compie fatti o atti diretti a sostituire denaro o valori proventi dai delitti di rapina aggravata, estorsione aggravata, o sequestro di persona a scopo di estorsione con altro denaro o altri valori, al fine di procurare a sé o ad altri un profitto o aiutare gli autori dei delitti suddetti ad assicurarsi il profitto del reato, è punito con la reclusione da quattro a dieci anni e con la multa da uno a venti milioni di lire.

Con questo decreto di legge viene per la prima volta identificato il reato di riciclaggio come indipendente dall'effettiva punibilità del crimine connesso (rapina aggravata, estorsione ecc).

Le numerose critiche che seguirono al decreto riguardarono la limitata applicazione dell'articolo, in quanto erano espressamente indicati i crimini connessi al riciclaggio.

Durante gli anni Ottanta aumentò la consapevolezza dei danni causati dal riciclaggio all'economia e nel 1989, al seguito del G7 tenutosi a Parigi, venne costituito il FATF (Financial Action Task Force). Il FATF introdusse 40 Raccomandazioni, ovvero una guida contenente una serie di contromisure per combattere il riciclaggio di denaro.

L'articolo 628bis venne modificato nel 1990 con l'introduzione dell'articolo 23, che ne estese l'applicazione anche al reato di *traffico di sostanze stupefacenti o psicotrope* e ne ampliò l'oggetto del reato (aggiungendo *beni e altra utilità*, così da colpire anche eventuali investimenti).

Fuori dai casi di concorso nel reato, chiunque sostituisce denaro, beni o altre utilità provenienti dai delitti di rapina aggravata, di estorsione aggravata, di sequestro di persona a scopo di estorsione, o dai delitti concernenti la produzione o il traffico di sostanze stupefacenti o psicotrope, con altro denaro, altri beni o altre utilità, ovvero ostacola l'identificazione della loro provenienza dai delitti suddetti, è punito con la reclusione da quattro a dodici anni e con la multa da Lire due milioni a Lire trenta milioni. La pena è aumentata quando il fatto è commesso nell'esercizio di una attività professionale. Si applica l'ultimo comma dell'articolo 648.

Un dettaglio interessante del nuovo 628bis riguarda la sostituzione della formula *compie fatti o atti diretti a sostituire* con *sostituisce*, che impone l'identificazione dell'effettiva sostituzione del denaro affinché sia riconosciuto il crimine.

Con la stessa legge viene introdotto anche l'articolo 628ter, che introduce un nuovo strumento per identificare e condannare il riciclaggio.

Chiunque, fuori dai casi di concorso nel reato e dei casi previsti dagli articoli 648 e 648-bis, impiega in attività economiche o finanziarie denaro, beni o altre utilità provenienti dai delitti di rapina aggravata, di estorsione aggravata, di sequestro di persona a scopo di estorsione, o dai delitti concernenti la produzione o il traffico di sostanze stupefacenti o psicotrope, è punito con la reclusione da quattro a dodici anni e con la multa da Lire due milioni a Lire trenta milioni. La pena è aumentata quando il fatto è commesso nell'esercizio di un'attività professionale. Si applica l'ultimo comma dell'articolo 648.

Nel 1991 la Comunità Europea emanò la Prima Direttiva Antiriciclaggio, la 91/308/CE, che definì, tra le altre cose, l'obbligo di registrazione delle operazioni per gli stati membri.

Possiamo identificare la base normativa per la lotta al riciclaggio in Italia nella legge 197, emanata nel 1991. Questa legge con 14 articoli definì nel dettaglio i precetti, i divieti, gli obblighi e le sanzioni, così come recepiti dalle Raccomandazioni del FATF. Con questa legge l'Italia si allineò con la Prima Direttiva: la legge vietò il trasferimento di denaro contante, di libretti di deposito bancari o postali al portatore o titoli al portatore tra soggetti privati se il valore dovesse superare i venti milioni di lire e, successivamente all'arrivo dell'euro, ai 12.500,00€. Per effettuare tali transazioni venne reso necessario rivolgersi a intermediari abilitati [Fal09].

L'articolo 23 venne nuovamente rivisto e corretto con la legge del 9 agosto 1993, ampliando ulteriormente l'insieme dei reati associati (fino a quel momento limitati alle quattro tipologie di crimini elencati) sostituendoli tramite la formula "provenienti da delitto non colposo" [Ber96].

Durante il 1997 le competenze e le responsabilità decisionali in materia di riciclaggio vennero trasferite all'UIC (Ufficio Italiano Cambi), limitando le responsabilità del ministero del Tesoro al coordinamento. Abbandonando l'iter proposto dalla legge del 1991, in cui l'intermediario comunicava direttamente con la Questura localmente competente, la nuova legge aumentò decisamente la cosiddetta *Collaborazione attiva* delle banche, decuplicando l'efficacia della lotta al riciclaggio.

Il 12 dicembre del 2001 venne realizzata dalla Comunità Europea la Seconda Direttiva antiriciclaggio che estese l'ambito di applicazione riguardo gli obblighi di identificazione alla clientela anche ai consulenti del lavoro, ai soggetti iscritti all'albo dei ragionieri, dei periti commerciali, nel registro dei revisori contabili e nell'albo dei dottori commercialisti. Inoltre assegnò obblighi di vigilanza anche a notai e avvocati. In Italia fu recepita solamente nel 2004 con il decreto legislativo 56/2004 [Vil08].

Verso la fine del 2005 venne emanata la Terza Direttiva antiriciclaggio, in cui si definiscono l'obbligo di adeguata verifica della clientela, anche per prestazioni occasionali, l'obbligo di registrazione delle informazioni acquisite e l'obbligo di segnalazione delle operazioni sospette all'Unità di Informazione Finanziaria (UIF) o all'Ordine Professionale di appartenenza.

Con il decreto 231 del 2007 l'Italia si adeguò alle disposizioni della direttiva. Nel testo venne, tra le altre cose, ridefinito un limite per il trasferimento del denaro contante, che non può essere pari o superiore a 5000,00€. Questo limite verrà modificato con il decreto del 2008 e nuovamente riportato a 5000,00 nel 2010 [Dip08].

Alla fine del 2009 vennero pubblicati nuovi provvedimenti antiriciclaggio. Oltre a far salire il limite oltre al quale viene richiesto l'obbligo di identificazione della clientela da 12.500,00€ a 15.000,00€ vengono emanate le nuove disposizioni sulla tenuta dell'Archivio Unico Informatico da parte della Banca d'Italia. Queste disposizioni sono entrate in vigore il 1 giugno 2010 e in questo momento sono le ultime disposizioni vigenti.

Da maggio del 2011 entreranno in vigore le nuove disposizioni per l'invio dei dati aggregati e verrà delineata la nuova figura del Responsabile Antiriciclaggio.

Capitolo 3

Lo stato dell'arte

3.1 L'Archivio Unico Informativo

L'Archivio Unico Informativo (AUI) è, come la parola *Archivio* lascia intendere, un archivio dove vengono registrate le transazioni bancarie. La parola *Unico* lascia erroneamente supporre che l'auì sia unico a livello italiano, mentre attualmente ogni banca ne possiede uno. Infine la parola *Informativo* definisce il supporto utilizzato per la memorizzazione dei dati.

La struttura dell'archivio è molto semplice, altro non è che un elenco di registrazioni con associati i soggetti coinvolti. Ad esempio un bonifico da un soggetto A ad un soggetto B verrà inserito nell'archivio con tre voci distinte ma collegate tra di loro. La prima rappresenterà la transazione, ovvero il bonifico in uscita. La seconda voce sarà la voce anagrafica del soggetto A, collegato alla voce del bonifico con il ruolo di soggetto che ha eseguito l'operazione. La terza voce, infine, sarà l'anagrafica del soggetto B, collegata al bonifico con il ruolo di soggetto controparte.

Esistono cinque tipologie di associazioni tra le anagrafiche e le operazioni, identificate con il nome di Tipi Informazione e indicizzati da due a sei. Per approfondimenti sulla numerazione e sulla struttura dei campi dell'archivio rimandiamo all'Appendice A.

Possiamo in ogni caso aggregare i campi dell'AUI in macro categorie. Ciascuna operazione registrata in AUI presenta campi riguardanti l'identificativo dell'informazione, la dipendenza interessata all'operazione, le caratteristiche dell'operazione, i dati sull'intermediario della controparte, i dati sul rapporto movimentato e i dati sull'importo. Per quanto riguarda l'anagrafica avremo anche qui l'identificativo dell'informazione, le complete generalità, la classificazione del soggetto, i dati sul rapporto movimentato e gli estremi del documento.

All'interno dell'archivio unico finiscono tutte le registrazioni che superano i 15.000,00€ singolarmente o nell'arco di 7 giorni di attività. Ad esempio se il 5 aprile il soggetto A effettua un bonifico per un totale di 15.000,00€ verso il soggetto

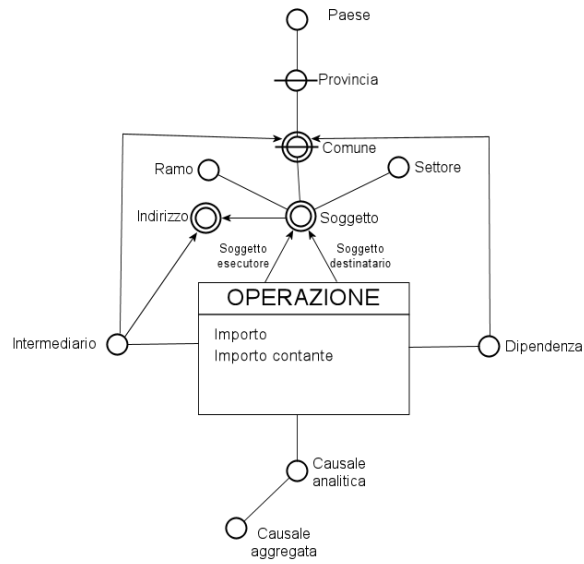


Figura 3.1: Datamart dell'AUI

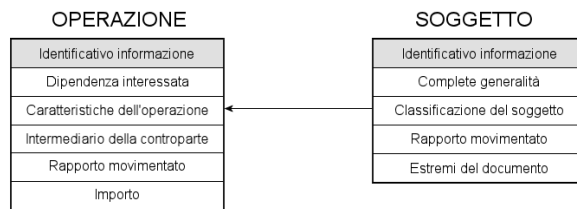


Figura 3.2: Informazioni presenti in una registrazione nell'AUI

B l'operazione finirà nell'AUI il 5 maggio. Se il soggetto A invece di effettuare un unico bonifico ne effettua cinque di importo 3.000,00€ l'uno, uno al giorno per cinque giorni dal 5 al 9 aprile, il 9 maggio finirà nell'AUI come operazione frazionata. Se infine il soggetto A effettuasse cinque bonifici da 3.000,00€ l'uno, uno ogni due giorni, all'interno della finestra di 7 giorni richiesta dalla normativa al massimo l'importo sarà pari a 12.000,00€, ovvero inferiore alla soglia, dunque le registrazioni non finiranno in AUI.

L'archivio viene generato giornalmente dagli intermediari analizzando le transazioni avvenute 30 giorni prima, ovvero un'operazione avvenuta il 6 novembre entrerà all'interno dell'AUI solamente il 6 di dicembre.

3.2 Il software Gianos®©

Il software GIANOS®© venne sviluppato nel 1993 al fine di aiutare e supportare l'operatore nell'adempimento delle sempre più complesse normative antiriciclaggio. Il progetto venne approvato dall'Associazione Bancaria Italiana, dall'Associazione Nazionale delle Banche Popolari e dall'Associazione fra Banche Ordinarie di Credito che avviarono un apposito gruppo di lavoro coordinato dall'ABI [1].

Dopo una lunga e delicata analisi, venne lanciata una prima versione in via sperimentale. Il 3 gennaio del 1995 si concluse positivamente il test sulle banche pilota e venne deciso di attivare il software su tutte le banche che avevano deciso di adottarlo. In breve tempo il programma si diffuse, arrivando a coprire circa il 97% delle dipendenze bancarie. Con l'evolversi della normativa GIANOS®© venne adattato e integrato per restare al passo con le disposizioni legali vigenti. Al momento l'applicativo viene realizzato e gestito da Oasi Diagram S.p.A. ed è giunto alla terza versione, realizzata in seguito dell'entrata in vigore della Terza Direttiva antiriciclaggio del 26 ottobre 2005.

3.3 La struttura di Gianos®©

GIANOS®©-3D (Terza Direttiva) è composto da tre moduli principali: il modulo Know Your Customer, il modulo per la Gestione dei Profili di Rischio (GPR) e il modulo per la Valutazione degli Inattesi (Fig. 3.3).

3.3.1 Gianos®© - Know Your Customer

Questo modulo gestisce ed elabora i questionari da rivolgere agli utenti nell'ambito dell'adeguata verifica ai sensi dell'art. 18 del D.lgs 231/2007. Il programma permette infatti di compilare l'apposito questionario richiesto dalle banche e di calcolare al termine il profilo di rischio iniziale del cliente, che verrà poi modificato dai successivi movimenti.

Le banche richiedono la compilazione del questionario non solamente per poter identificare l'utente, ma anche per conoscerne le intenzioni e i comportamenti. In

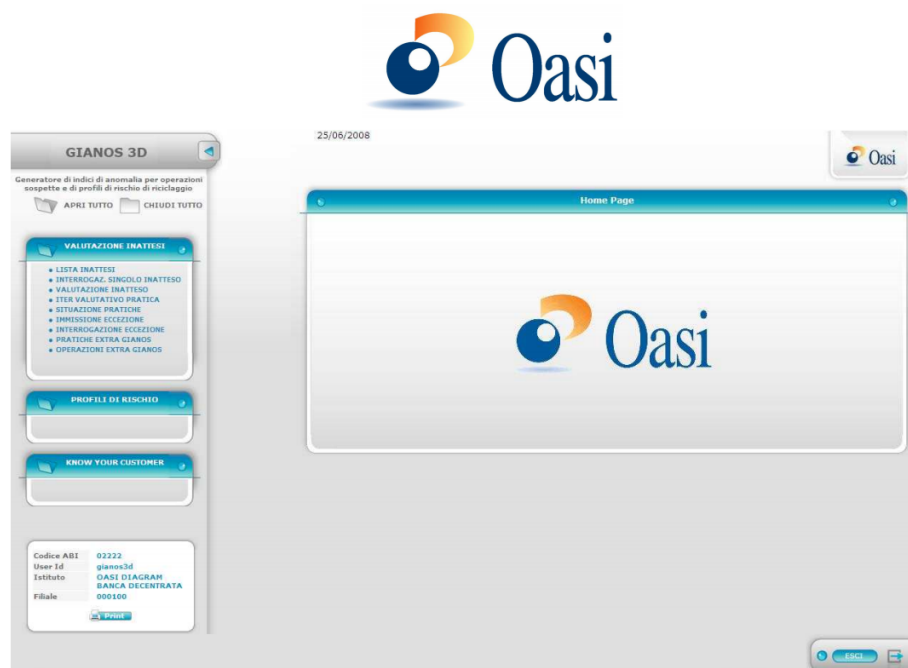


Figura 3.3: Il programma GIANOS®©

questo modo, se il soggetto ad esempio dichiara di non voler effettuare investimenti rischiosi e poi disattende questo suo proposito, il comportamento sarà sospetto e potrebbe influenzare il punteggio associato.

In generale possiamo sintetizzare le informazioni del questionario in due macro categorie: le informazioni sul cliente e le informazioni sulle attività del cliente. Nella prima categoria saranno presenti domande riguardanti i dati del cliente (nome, cognome, codice fiscale) i dati del titolare del rapporto (se diverso dal cliente), i dati sulla residenza e il documento presentato. Nella seconda categoria invece saranno indicati i dati del rapporto, ovvero il settore (SAE e RAE) e le informazioni riguardanti la natura e lo scopo.

Inoltre il soggetto può venire associato ad una serie di liste di sorveglianza, come le liste terroristiche o la lista delle persone politicamente esposte (Politically Exposed Persons).

3.3.2 Gianos®© - Gestione Profili di Rischio

Il modulo GPR contiene le procedure per il calcolo del profilo di rischio del soggetto. Analizzando infatti i dati del cliente, lo storico delle sue transazioni e il numero e la tipologia di rapporti che il cliente ha avviato con l'intermediario il programma elabora e modifica il punteggio associato al soggetto, variandone eventualmente la fascia di rischio (Fig. 3.4).

Dettaglio Del Punteggio Calcolato

• Soggetto :

• Cognome/Ragione Sociale :

• Punteggio e Fascia Iniziale : 001 - IRRILEVANTE

• Punteggio e Fascia Attuali : 053 - ALTA

Pagina : 1 / 2 ◀ precedente | successivo ▶

Componente	Punti	segno	Componente	Punti	segno
OPERAZIONI DARE	07	+	OPERAZIONI IN CONTANTE	04	+
PUNTEGGIO ECCEZIONI	02	-	PUNTEGGIO INATTESI	03	+
PUNTEGGIO EXTRAGIANOS	02	+	CLIENTE AFFIDATO	01	-
ASSEGNI LIBERI	01	+	CANALE INTERNET/TELEFONICO	03	+
NUMERO CARTE PREPAGATE	01	+	NUMERO CASSETTE SICUREZZA	01	+
DELEGATO OPERARE SU + CONTI	01	+	GIACENZA MEDIA	02	+
NUMERO CONTI DISPON.OPERARE	01	+	NUMERO OPERAZIONI	03	+
OPERATIVITA'	03	+	PAESE DI RESIDENZA	02	+
STRUMENTI FINANZIARI	01	+	PUNTEGGIO CONDIZIONI 002	12	+
PUNTEGGIO CONDIZIONI 008	03	+	PUNTEGGIO CONDIZIONI 010	01	+

Stampa Report nel formato:

Inizio **Lista** **Indietro**

Figura 3.4: Dettagli del profilo di rischio di un soggetto

Questi punteggi vengono determinati da regole personalizzabili volte a profilare determinati insiemi di circostanze. Ad esempio potrebbe venire associato un punteggio alle persone con più di settant'anni delegate ad operare in più di dieci conti o soggetti con Settore o Ramo di Attività Economica sensibile all'utilizzo del contante (Fig. 3.5).

Lista Associazione Punteggio Condizione

Pagina : 2 / 4 ◀ precedente | successivo ▶ Num. Punteggio 31

Scelta	Condizione	Descrizione	Punti	Data Inizio	Data Fine
<input type="radio"/>	013	SOGGETTI CON ETA' SUPERIORE A 70 ANNI CHE	008	01/07/2007	01/07/2099
<input type="radio"/>	014	SOGGETTO DELEGATO CON + 70 ANNI E RESIDEN	002	01/01/2008	31/12/2099
<input type="radio"/>	015	SOGGETTO DELEGATO CON + 70 ANNI CON ATTIV	001	01/02/2008	31/12/2099
<input type="radio"/>	016	IMPORTO OPERAZIONE > 5000,00 < 15000,00	001	01/02/2008	31/12/2099
<input checked="" type="radio"/>	017	IMPORTO OPERAZIONE > 15000,00 < 30000,00	002	01/02/2008	31/12/2099
<input type="radio"/>	018	IMPORTO OPERAZIONE > 30000,00	003	01/02/2008	31/12/2099
<input type="radio"/>	019	TITOLARE EFFETTIVO	001	01/02/2008	31/12/2099
<input type="radio"/>	020	SE DIRIGENTE/IMPREDITORE/LIBERO PROFES.	008	01/01/2008	31/12/2099

Stampa Report nel formato:

Variazione Elemento **Dettaglio Elemento**

Indietro

Figura 3.5: I punteggi associati ad alcune regole

Ogni volta che si verificherà uno di questi casi il punteggio del soggetto cambierà conformemente alla variazione in punti associata all'evento (Fig. 3.6).

Codice Soggetto	Denominazione	Puntti Attuale	Fascia	Puntti Iniziale	Filiale
000020000201166F	AZIENDA PIPPO	87	ALTA	00	000020
000020000200400F	AZIENDA PROVA	05	IRRIL	00	000020
000020000200327F	BIANCHI SANDRA	05	IRRIL	00	000020
000020000200398F	DITTA	09	MEDIA	00	000020
000020000201575F	GIALLI FRANCO	05	IRRIL	00	000020
000020000200204G	PROVA PROVA	63	ALTA	00	000020
000020000201056F	ROSSI CARLO	04	IRRIL	00	000020
000020000200510F	SOCIETA'	05	IRRIL	00	000020
000020000201469F	SOCIETA' PAPERINO	55	ALTA	00	000020
000020000201248F	SOCIETA' PLUTO	05	IRRIL	00	000020

Figura 3.6: Variazione dei profili di rischio

3.3.3 Gianos®© - Valutazione Inattesi

La valutazione degli inattesi rappresenta il cuore del programma GIANOS®©. Tramite questo modulo, infatti, è possibile identificare le transazioni sospette (inattesi GIANOS®©) di riciclaggio. La selezione degli inattesi avviene tramite algoritmi segreti definiti dal Comitato ARMA (AntiRiciclaggio Monitoraggio Applicazione Procedura GIANOS®©), un gruppo di lavoro interbancario costituito dall'ABI. Gli algoritmi sono costantemente monitorati tramite statistiche ed eventualmente modificati dallo stesso Comitato.

GIANOS®© analizza mensilmente i dati dell'Archivio Unico Infomatico relativi alle operazioni verificatesi due mesi prima del mese corrente. All'interno delle operazioni GIANOS®© utilizza gli algoritmi segreti per individuare le transazioni sospettate di riciclaggio. Il software offre anche la possibilità di aggiungere *eccezioni*, ovvero di segnalare determinati inattesi come *inattesi soppressi*, non destinati a valutazione. In ogni caso resta possibile esaminare queste segnalazione tramite ricerca esplicita e verificare chi ha creato l'eccezione e le motivazioni.

Una volta individuati è possibile richiamare la lista e controllare i dettagli di ogni inatteso (Fig. 3.7). Tra le informazioni disponibili, oltre ai dati del soggetto e ai dettagli della transazione (Fig. 3.8), è possibile visualizzare i motivi che hanno fatto partire la segnalazione dell'inatteso (Fig. 3.9).

3.4 Il funzionamento di Gianos®©

GIANOS®© riceve in input l'AUI integrato con alcuni dati aggiuntivi personalizzati a seconda delle necessità dell'intermediario. Questi dati aggiuntivi compongono l'AUI Complementare Esteso (AUICE). Una volta verificati i dati in input, il programma elabora le registrazioni presenti in AUI e genera i profili per tutta la clientela che abbia effettuato almeno una registrazione negli ultimi 12 mesi. Qua-

Lista Inattesi Selettiva

◀ precedente | successivo ▶ Numero inattesi 16674

Codice Soggetto	Denominazione	Periodo Riferim	Stato	Filiale Incaric	Capo Gruppo	Tipo Pratica	Pratiche Soppresse
<input type="radio"/> 000020000200398F		2007 01	01			E	N
<input type="radio"/> 000020000200398F		2008 07	00			E	N
<input type="radio"/> 000020000200580F		2005 10	00			G	N
<input type="radio"/> 000020000200580F		2005 10	03	000020	000020	G	N
<input type="radio"/> 000020000201806G		2006 10	00			G	N
<input type="radio"/> 000020000201806G		2006 10	03	000020	000020	G	N
<input type="radio"/> 000020000201806G		2008 07	00			E	N

Stampa Report nel formato: [Icone]

Versione 02,007

Figura 3.7: Lista degli inattesi

Interrogazione Operazioni Inatteso

• Cod. Soggetto: • Stato:

• Descr Soggetto: • Num. Oper. :

• Data Op. • Dipendenza

• Id. Registr • Divisa

• Segno • Paese

• Imp. Operaz. • Imp. Contan.

• Ruolo • Cod. Rapporto

• Cod. Causale • Causale

• A disposizione

• Data Op. • Dipendenza

• Id. Registr • Divisa

• Segno • Paese

• Imp. Operaz. • Imp. Contan.

• Ruolo • Cod. Rapporto

• Cod. Causale • Causale

• A disposizione

Versione 01,008

Figura 3.8: Dati di un inatteso



Figura 3.9: Dettaglio di un inatteso

lora fossero disponibili le informazioni nell'AUICE, GIANOS®© creerà i profili anche per i soggetti che non dovessero ricadere in questo insieme.

3.5 I comportamenti

Allo stato attuale GIANOS®©, tramite gli algoritmi secretati, cerca di prevedere e di anticipare le strategie, gli atteggiamenti e le attività comuni tra i riciclatori di denaro. Questa forma di analisi è molto delicata e chiaramente tende ad essere sbilanciata verso i falsi positivi: meglio una falsa segnalazione di un crimine impunito.

Gli algoritmi individuano una serie di comportamenti standard, elencati nell'Appendice B in base alle disposizioni emanate dalla Banca d'Italia. Nell'elenco, per comprendere le categorie, sono presenti voci come *Aziende in crisi con movimenti in avere nel mese* o *Pareggiamento tra avere e dare*. La prima categoria dell'esempio identifica le aziende in crisi che improvvisamente ricevono un ingresso di denaro mentre la seconda identifica dei conti utilizzati solamente come *tramite* per trasferimenti.

Come si può vedere dall'appendice, spesso si tratta di comportamenti complessi che, per poter venire correttamente identificati, richiedono algoritmi altrettanto complessi.

Capitolo 4

Un po' di matematica

L'analisi a posteriori di questo studio cerca di comprendere, una volta ricevuta la segnalazione di un inatteso da parte di GIANOS®©, quanto quell'inatteso è effettivamente anomalo. Quanto è *inattesa* quella transazione per quel determinato soggetto. L'idea è identificare delle anomalie comportamentali nell'individuo *dopo* la segnalazione di GIANOS®©, così da facilitare il successivo compito dei valutatori.

Per effettuare uno studio in tal senso, bisogna premettere alcune nozioni utilizzate.

4.1 Le ciclicità nel comportamento

L'uomo è abitudinario, tende a ripetere le azioni che maggiormente lo porteranno ad avere *successo*. Quindi se acquista dal supermercato sotto casa e non ne rimane deluso, tenderà a tornarci una seconda volta, poi una terza e così via.

Ma questa forma di abitudine spesso agisce anche temporalmente. Se il venerdì non trovo molta fila, tenderò a tornare il venerdì.

A questo si aggiungono anche altre forme di ciclicità, come la retribuzione mensile di un dipendente o il pagamento dell'affitto di un appartamento.

Similmente possiamo considerare i comportamenti aziendali come ciclici. Anche qui avremo il pagamento dei salari, eventuale pagamento dell'affitto degli immobili o pagamento delle tasse.

Ovviamente all'interno della vita di un individuo o di un'azienda avverranno dei cambiamenti. I cicli tenderanno a mutare in base agli eventi e alle scelte effettuate dal soggetto considerato. Per questo motivo occorrerà limitare lo storico considerato per l'individuazione del ciclo, in modo da renderlo adattabile ai cambiamenti ed evitare un appiattimento delle reazioni alle modifiche.

Possiamo dunque considerare anomala una qualsiasi operazione che non ricada negli atteggiamenti standard dell'individuo o nell'azienda, sia come importo che come periodo.

Supponiamo ad esempio che un impiegato guadagni 1.800,00€ il 27 di ogni mese, paghi regolarmente un affitto di 600,00€ il 5 del mese e spenda 250,00€ a settimana di spese varie.

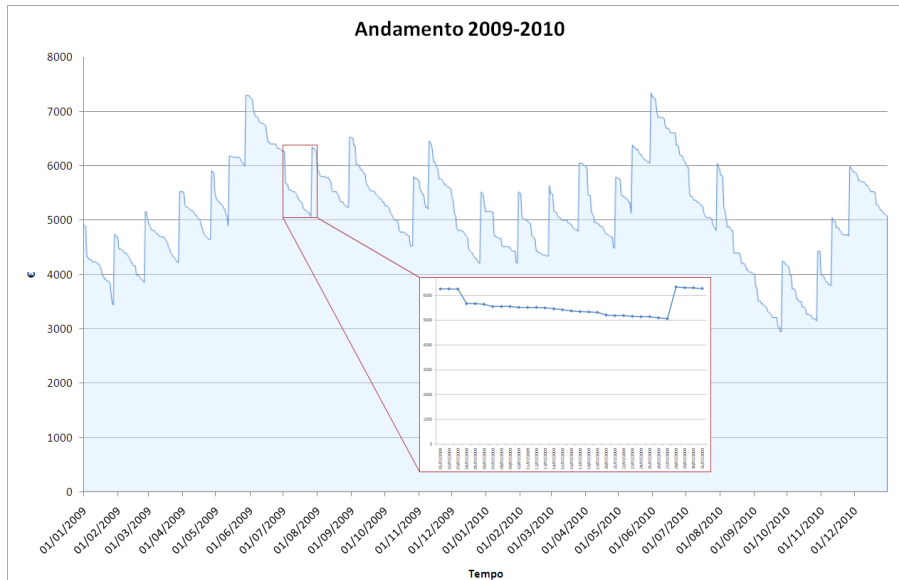


Figura 4.1: Grafico tipo di un dipendente

Ovviamente se improvvisamente ricevesse un milione di euro sarebbe un inatteso. Ma se dovesse ricevere 1.800,00€ il 15 del mese o pagasse 600,00€ il 20 del mese sarebbero comunque inattesi, anche se più difficilmente rilevabili.

A questo scopo dobbiamo analizzare il comportamento dell'utente estraendone le informazioni attinenti dallo storico delle transazioni.

4.2 Anomalia dell'importo

Ai fini dell'individuazione di un'anomalia dobbiamo prima di tutto analizzare il comportamento *standard* degli intermediari durante un determinato ciclo.

Il campione di dati utilizzato è stato estratto da un Archivio Unico presente in una banca e debitamente anonimizzato per poter essere utilizzato a scopi statistici.

Dai dati sono state estratte le transazioni, suddivise per soggetto esecutore/titolare e per data e sono state effettuate le relative somme del dare e dell'avere in un determinato giorno.

I dati raccolti e suddivisi per soggetto sono stati esaminati per valutare la tipologia di distribuzione presente nelle ciclicità dei singoli soggetti. Per fare questo è stato scelto di volta in volta un giorno del mese e sono stati analizzati gli eventi appartenenti al periodo di due giorni precedenti e successivi alla data scelta.

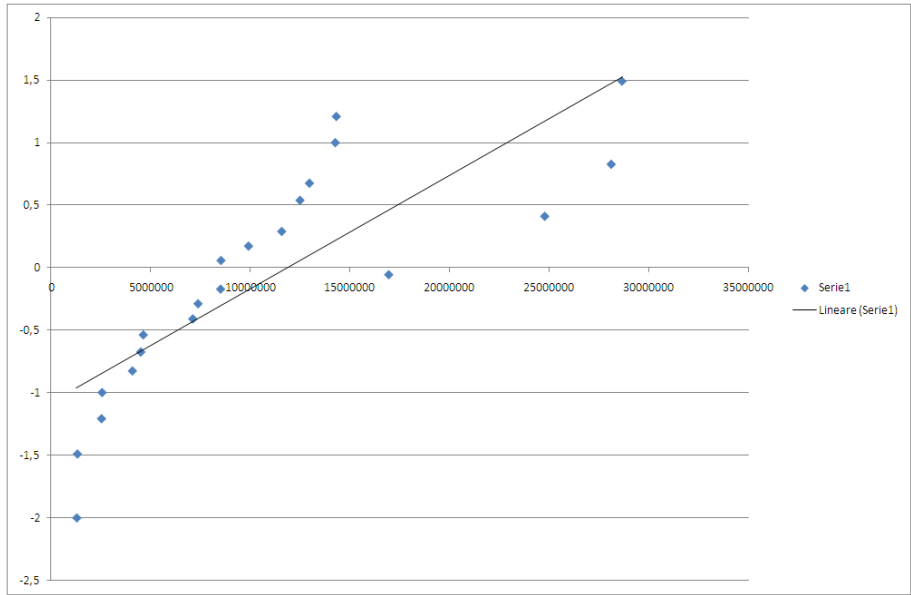


Figura 4.2: Distribuzione normale al 15 del mese

Ad esempio se il giorno scelto fosse stato il 24, sarebbero state prese tutte le transazioni con data compresa tra il 22 e il 26 di ciascun mese.

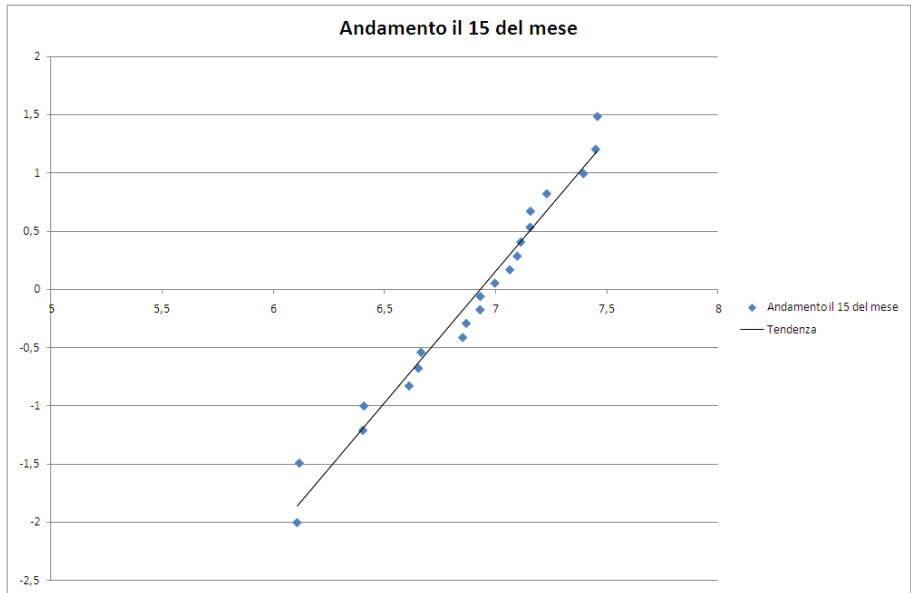


Figura 4.3: Distribuzione lognormale al 15 del mese

Una tecnica semplice ed efficace per identificare se una distribuzione ha una distribuzione normale consiste nel cosiddetto *Normal Probability Plot*, in cui la

successione di numeri da valutare viene ordinata e rappresenta graficamente in diretta correlazione con distribuzione normale. In questo modo, se i punti dovessero avere una distribuzione normale, andrebbero a disporsi linearmente.

Le prime prove effettuate erano volte ad individuare se la distribuzione di interesse avesse una distribuzione normale (Fig 4.2) o meno. Analizzando il risultato è stato concluso che la distribuzione non aveva caratteristiche normali. A questo punto è stata esaminata la possibilità che la distribuzione fosse lognormale.

Una variabile casuale X ha una distribuzione lognormale con parametri μ e d se il suo logaritmo ha una distribuzione normale μ con deviazione standard d . Nel nostro caso dunque è bastato effettuare il logaritmo ed eseguire il Normal Probability Plot sul risultato. Il grafico è mostrato in figura 4.3 citeLognormal09. Come si può facilmente osservare, i punti si allineano lungo la linea di tendenza, indicando quindi che la distribuzione di queste ciclicità è probabilmente lognormale.

4.3 Il test statistico

Per poter affermare che una determinata osservazione fa o meno parte di un campione è stato utilizzato il test statistico. Questo test altro non è che un'analisi delle caratteristiche del campione. Una volta esaminato l'andamento del campione viene effettuato il calcolo per poter valutare se è realistico o meno che si sia verificato il caso osservato.

Nel caso in questione ovviamente il dato risulta interessante proprio se risulta improbabile nella distribuzione del campione, ovvero se è un'anomalia comportamentale dell'individuo. Essendo presenti poche osservazioni (nel caso migliore di una ciclicità mensile avremo al massimo 60 osservazioni l'anno, ma in media sono circa 20 l'anno) utilizzeremo un test di Student [Sco06] [Stu09].

$$t = \frac{\bar{X} - \mu_0}{\sigma \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

Se ad esempio abbiamo un campione di 20 osservazioni con

$$\bar{X} = 2500$$

$$\sigma = 300$$

mentre l'osservazione è pari a

$$\mu_0 = 2000$$

avremo che

$$t = \frac{2600 - 2000}{300 \sqrt{\frac{1}{20} + 1}}$$

$$t = \frac{600}{300\sqrt{\frac{21}{20}}}$$

$$t = 1,951800146$$

Confrontandolo con un test al 0,05 e 20 gradi di libertà avremo che

$$1,721 < 1,951800146$$

Dunque possiamo rifiutare l'ipotesi nulla $H_0 = \mu > \mu_0$.

Nel nostro caso non andremo ovviamente a confrontare il valore della nuova osservazione, ma il logaritmo del valore, con il risultato del test. Il risultato finale sarà visualizzato all'utente come un valore nominale composto da cinque possibili valori, ottenuti confrontando il risultato del test con i valori di alpha al 0,1 allo 0,05 e allo 0,01.

4.4 Anomalia del periodo

Come affermato in precedenza dobbiamo considerare anomalo non solo un importo quantitativamente straordinario, ma anche un importo temporalmente fuori dall'ordinario.

Per controllare se un'operazione risulta temporalmente fuori dall'ordinario basta controllare la frequenza media e la varianza del periodo di interesse, considerando separatamente i vari cicli possibili.

Ad esempio per controllare se una data transazione avvenuta il 15 gennaio è anomala temporalmente considereremo l'intervallo dal 13 al 17 di ciascun mese, controllando con quale frequenza le operazioni si presentano in quel periodo. Inoltre effettueremo gli stessi calcoli per i trimestri e i semestri (a partire da gennaio). Infine verrà valutata anche la frequenza annuale.

La formula sarà dunque

$$0.4 \cdot \frac{N_{ok}^d}{N_{tot}^d} + 0.2 \cdot \frac{N_{ok}^{d+1}}{N_{tot}^{d+1}} + 0.2 \cdot \frac{N_{ok}^{d-1}}{N_{tot}^{d-1}} + 0.1 \cdot \frac{N_{ok}^{d+2}}{N_{tot}^{d+2}} + 0.1 \cdot \frac{N_{ok}^{d-2}}{N_{tot}^{d-2}}$$

Dove N_{ok}^d indica il numero delle transazioni avvenute durante il giorno considerato mentre N_{tot}^d sono le transazioni totali. I giorni appartenenti all'intorno della data scelta vengono inseriti con peso minore nel calcolo. Il valore risultante viene anche in questo caso visualizzato all'utente come valore nominale distinto in cinque livelli di anomalia.

4.5 Pesare un inatteso Gianos®©

Il peso dell'inatteso è ottenuto tramite i due indicatori Anomalia nell'importo e Anomalia nel periodo. Questi due indicatori verranno estratti ogni qual volta verrà

segnalato da parte del programma un inatteso. Nel dettaglio il primo indicatore viene calcolato dal risultato del test di Student sull'importo, mentre il secondo indicatore viene rappresentato dalla probabilità del verificarsi della transazione in quel determinato momento. Entrambi gli indicatori sono rappresentati da un intero compreso tra 1 e 5.

I due indicatori vengono rappresentati a video come separati e serviranno come informazione aggiuntiva all'operatore. Inizialmente non saranno unificate le due probabilità, in quanto è sconosciuta la formula di composizione.

4.6 Analisi dei risultati

Ricapitolando il processo di analisi abbiamo che al verificarsi dell'inatteso il programma calcola due stimatori di anomalia, l'Anomalia dell'importo e l'Anomalia del periodo. L'operatore esaminerà la transazione tenendo conto delle due informazioni separatamente, valutando se l'operazione risulta sospetta o meno. Nel momento in cui prenderà la decisione l'inatteso verrà segnalato o meno come sospetto.

Sono necessari dei feedback sull'efficacia delle statistiche estratte, così da poter individuare subito eventuali imprecisioni nei comportamenti degli algoritmi.

Celato all'operatore, così da non influenzarne il giudizio, verranno calcolati i coefficienti per la composizione dei due indici. Per fare questo calcolo memorizzeremo le scelte dell'operatore per ciascuna operazione sospetta.

La formula di cui si desidera cercare i valori è:

$$X^a \cdot Y^b = v$$

che, trasformata in logaritmi, corrisponde a

$$a \cdot \log(X) + b \cdot \log(Y) = \log(v)$$

Dove X e Y sono i valori dei due indici, mentre a e b sono i parametri con cui verrà pesata la relazione. Il risultato, v , rappresenta un indice che verrà messo in relazione con la probabilità dell'evento.

Possiamo quindi utilizzare la regressione logistica.

L'indicatore s_1 di soglia verrà stimato tramite il predittore *logit*

$$s_0 = \frac{\exp(s_1)}{1 + \exp(s_1)}$$

Dove s_0 rappresenta la soglia da definire per valutare se la transazione risulta anomala. Una volta calcolato s_1 dobbiamo vedere se

$$X^a \cdot Y^b = v \geq s_1$$

Se questo dovesse risultare vero allora l'operazione risulterà anomala.

Per poter valutare quanto il predittore è funzionante verranno raccolte statistiche sul numero dei falsi positivi e dei falsi negativi, lasciando all'amministratore il compito di poter variare conseguentemente la soglia [Cia08].

Capitolo 5

Gli strumenti per l'analisi: SAS e Pentaho

Le analisi riportate in questo lavoro ci hanno permesso di provare due suite di Business Intelligence: SAS, sviluppato da SAS institute, e Pentaho Community Edition. Queste due soluzioni software rappresentano due dei principali tool di Business Intelligence esistenti sul mercato e differiscono essenzialmente per la licenza con cui vengono distribuiti. Pentaho Community Edition infatti è una suite Open Source, mentre SAS è un prodotto a licenza privata.

5.1 SAS

SAS è una potente suite di Business Intelligence che vanta un'esperienza trentennale nel campo.

L'elemento principale che compone la suite è SAS BI Server, il quale fornisce le funzionalità base come ad esempio la creazione di report o l'integrazione con i principali database. Eventuali funzioni specifiche sono disponibili all'interno di package installabili: ad esempio SAS/STAT contiene le funzioni statistiche.

SAS al suo interno utilizza un linguaggio proprietario composto essenzialmente di due operazioni: l'operazione *data* e l'operazione *proc*. La prima operazione consente di gestire i dati. Si occupa di leggere, interpretare, scrivere ed elaborare i *dataset* SAS, ovvero gli insiemi di dati resi accessibili all'interno del programma. La seconda operazione esegue le funzioni e le procedure sui dataset.

Esistono altre tipologie di operazioni, ma rappresentano istruzioni di limitata importanza relativamente alle due sopra esaminate. Ad esempio la funzione *libname*, utilizzata per creare dataset permanenti, o *goption*, utilizzato all'interno del pacchetto SAS/GRAPH per impostare le opzioni grafiche.

Un programma SAS non è altro che un susseguirsi di *data* e di *proc*. Per questo lavoro, prima dell'integrazione tramite Java, abbiamo sviluppato il programma SAS finalizzato a testare la validità dei ragionamenti.

5.2 Pentaho

La suite Pentaho è composta da numerosi programmi, ciascuno delegato ad operare in una specifica area della Business Intelligence. Tra i principali componenti a disposizione abbiamo Kettle, a cui sono delegati i processi di ETL (Extract Transform Load), Reporting, per la formattazione degli output, Mondrian, un server OLAP, e Weka, un tool per il Data Mining.

Essendo Open Source, Pentaho è a diretto contatto con il pubblico tramite una vasta community. Il forum presente sul sito è costantemente controllato da decine di sviluppatori in grado di fornire risposta a dubbi e problemi. La documentazione è vasta e completa, e sono disponibili numerosi tutorial per i meno esperti.

La suite è sviluppata integrando prodotti molto diversi, ma tutti sono sviluppati per avere un'interfaccia intuitiva e user friendly: difficilmente sono presenti operazioni che richiedono obbligatoriamente l'utilizzo della tastiera.

5.3 L'integrazione

Visto che entrambi i prodotti garantiscono l'integrazione con Java, all'interno del programma sviluppato è stata creata un'interfaccia al fine di rendere trasparente e intercambiabile l'utilizzo delle due suite. In questo modo è stato possibile effettuare l'analisi dei due tool per un eventuale utilizzo futuro all'interno del prodotto.

Per la valutazione sono stati utilizzati criteri soggettivi, basati sull'esperienza diretta di sviluppo nell'applicazione di funzionalità specifiche.

Capitolo 6

Lo sviluppo dell'applicazione

Durante lo svolgimento della tesi è stato realizzato un applicativo stand-alone che racchiudesse tutte le funzionalità necessarie per lo sviluppo del plugin per GIANOS®©. Il programma si articola nelle fasi elencate in Figura 6.1.

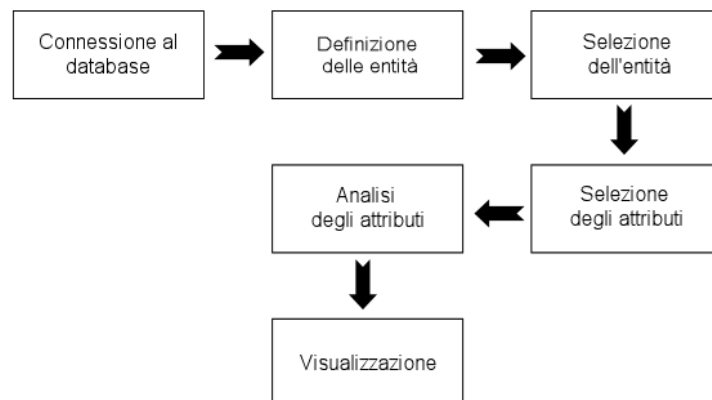


Figura 6.1: Fasi del programma

Le varie fasi hanno richiesto, principalmente per l'analisi, l'utilizzo dei software di Business Intelligence precedentemente segnalati.

6.1 Connessione al database

La prima schermata del programma consente di gestire le connessioni verso il database. L'idea è stata fornire un'interfaccia indipendente. Il menu infatti consente di definire il driver utilizzato, la stringa di connessione, il nome e la password (vedi Fig. 6.2).

Le connessioni vengono memorizzate su file xml, il che consente di gestire facilmente gli accessi a database diversi.

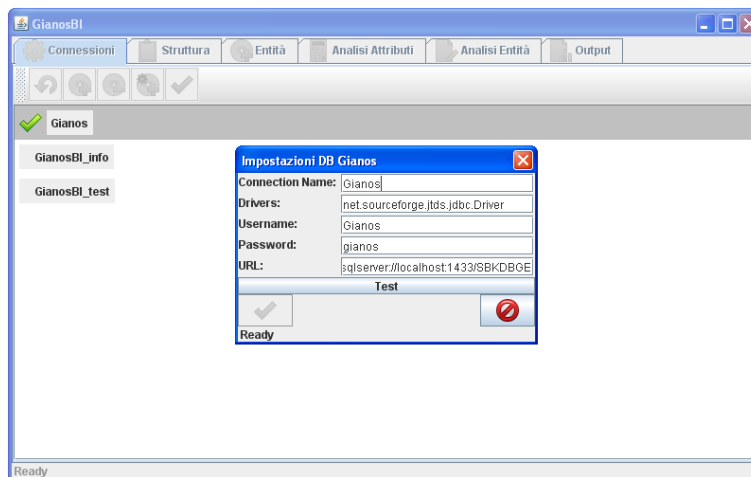


Figura 6.2: La creazione di una connessione

In questa fase viene definito anche il database che verrà utilizzato nelle analisi.

6.2 Le entità

Una volta connessi al database, andremo a definire degli aggregati di attributi su cui vogliamo effettuare l'analisi. A questo scopo la seconda schermata del programma consente di definire degli oggetti chiamati *entità*, che rappresentano una "vista" sulle tabelle.

Dopo aver definito le tabelle che si desidera interrogare e come effettuare i *join* si definiscono gli attributi che si desidera raggruppare in un'entità, e le si da un nome univoco (Fig. 6.3).

È possibile creare tutte le entità che si desidera mettere a disposizione per l'analisi. Anche le entità vengono salvate su xml, in modo da garantirne il riutilizzo.

L'elenco delle entità a disposizione viene visualizzato nella schermata *Entità* (Fig. 6.4).

Premendo su un'entità viene visualizzato l'elenco degli attributi e il relativo tipo di dato (numerico o testo). Da questa schermata è possibile visualizzare un primo insieme di statistiche sull'attributo (minimo, massimo, media, varianza, numero di elementi e valori distinti) così da avere un'idea del dato che si sta per analizzare.

6.3 L'analisi dell'entità

Una volta decisa l'entità che si desidera analizzare si passa alla schermata successiva, in cui è possibile esaminare nel dettaglio gli attributi numerici e visualizzarne le distribuzioni. In questo tab sarà possibile selezionare un sottoinsieme di attributi

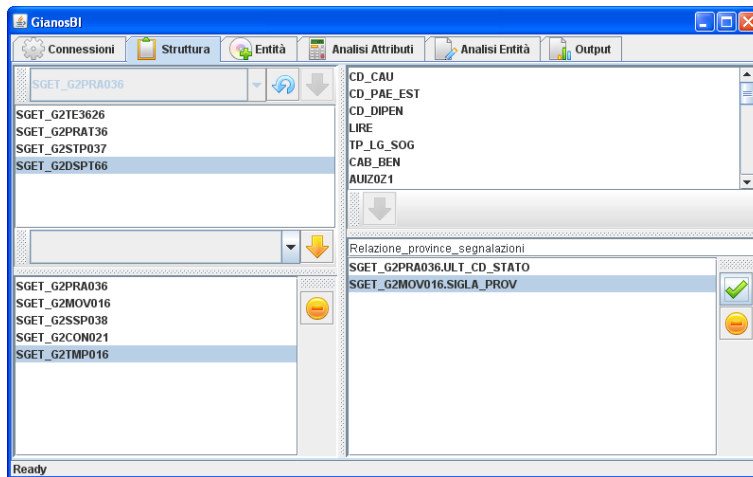


Figura 6.3: Definizione dell'entità di analisi

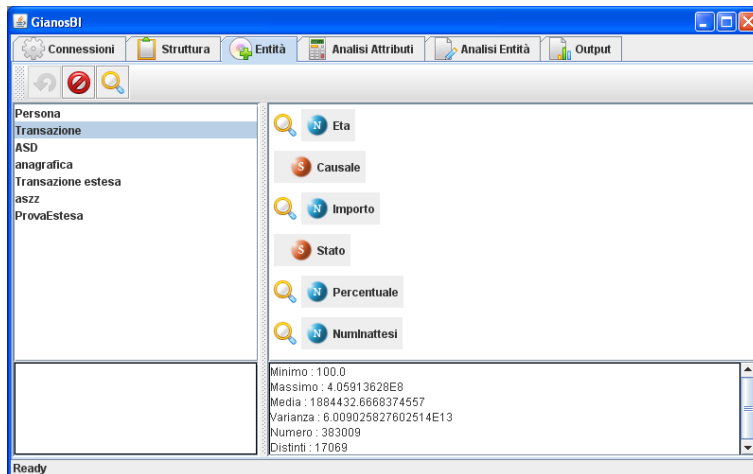


Figura 6.4: Elenco delle entità

e filtrarli in base alle necessità. Nell'esempio in Figura 6.5 possiamo vedere come l'età è stata filtrata tra i 35 e i 70 anni.

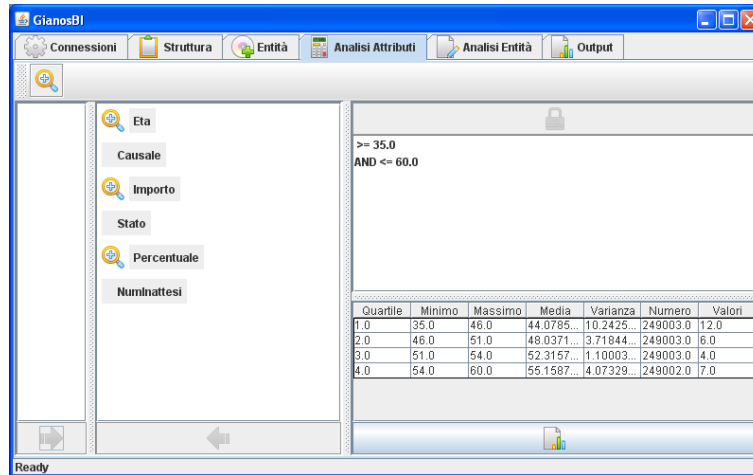


Figura 6.5: Analisi degli attributi numerici

L'insieme degli attributi selezionati viene passato alla schermata successiva, ovvero la schermata di analisi dell'attributo nominale.

Il tab *Analisi Entità* presenta l'elenco degli attributi nominali e, selezionandone uno, ne consente un'analisi approfondita (Fig. 6.6). La selezione dell'attributo nominale permette non solo di osservarne le distribuzioni, ma anche di analizzarlo come attributo-target delle varie analisi inserite nel programma (alberi decisionali, regressioni lineari, regressioni logistiche ecc.). Queste analisi sono state inserite per poter rapidamente visualizzare se un attributo è dipendente o meno da altri e in che modo.

Tramite questo tab è inoltre possibile visualizzare l'elenco delle anomalie trovate, eventualmente filtrate in base alle regole definite nella schermata precedente. La visualizzazione delle anomalie permette di esaminarne i due indicatori associati (Anomalia nell'importo e Anomalia nel periodo) definiti in precedenza (Fig. 6.7).

6.4 Creazione dell'output

Una volta analizzata l'entità, l'ultimo tab consente l'esportazione in pdf dei risultati ottenuti. Nel pdf vengono inserite tutte le statistiche effettuate, le analisi degli attributi e degli inattesi. Inoltre vengono calcolate alcune analisi ulteriori, come la regressione lineare e gli alberi decisionali utilizzando come valore target l'attributo selezionato nel tab precedente e come variabili i restanti attributi inseriti nell'analisi (Fig. 6.8).

In questo modo è possibile visualizzare eventuali correlazioni nascoste tra gli attributi.

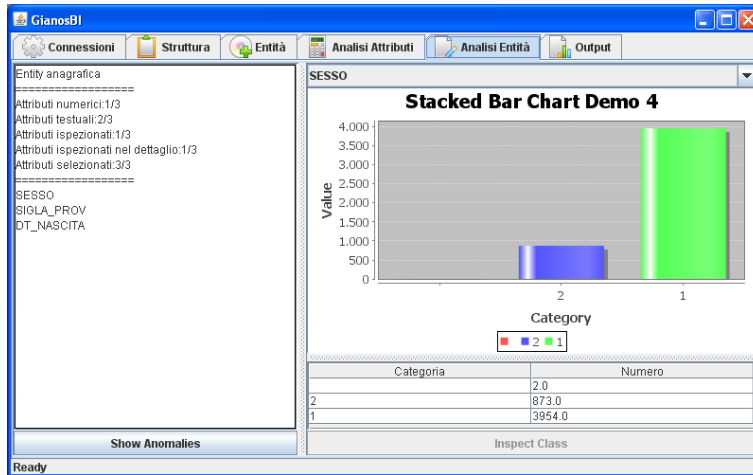


Figura 6.6: Analisi degli attributi nominali

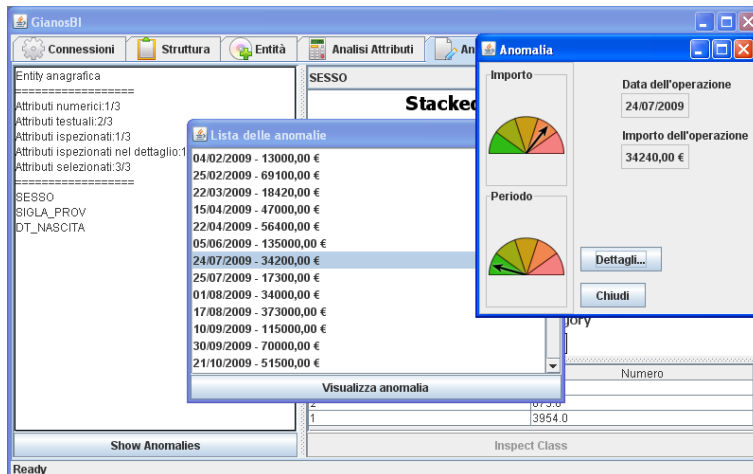


Figura 6.7: Analisi degli inattesi

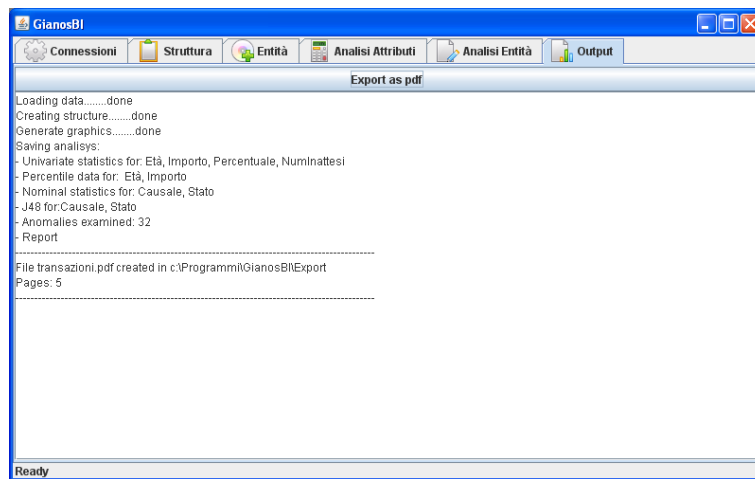


Figura 6.8: Analisi degli inattesi

Capitolo 7

Conclusioni

Il progetto è servito a individuare alcune aree di applicazione della Business Intelligence che potrebbero portare vantaggi nel mondo dell'antiriciclaggio. Vantaggi non solamente legati all'individuazione delle transazioni maggiormente sospette, ma anche al miglioramento del compito degli operatori e dunque all'ottimizzazione delle risorse aziendali.

La versione Community Edition di Pentaho fornisce ottimi strumenti, facilmente integrabili ma decisamente troppo legati all'interfaccia. L'astrazione spesso eccessiva impedisce di effettuare operazioni semplici per mancanza degli strumenti necessari. Pentaho consente ovviamente di implementare i propri strumenti, ma il procedimento è spesso troppo oneroso, risultando maggiormente fruibile l'utilizzo di linguaggi alternativi come java o javascript integrati all'interno del progetto.

SAS utilizza un linguaggio estremamente potente e risulta di molto maggiore come velocità di calcolo. Gli strumenti forniti nella versione Base SAS risultano completi e non vincolanti. Le estensioni del linguaggio permettono di ottenere vantaggi specifici in una determinata area, come il reporting o il data mining. La comunicazione con il java risulta piuttosto semplice per integrazioni non complesse, ma tende a complicarsi con l'aumentare del dettaglio delle richieste.

Per gli sviluppi successivi del progetto è stato deciso di adottare SAS in quanto maggiormente in linea con le necessità implementative future.

Capitolo 8

Sviluppi futuri

L'implementazione del test per l'identificazione degli inattesi è solo una delle tecniche che stanno venendo implementate per facilitare il compito di individuazione delle operazioni sospette. In particolare i possibili sviluppi futuri riguarderanno:

1. Implementazione dell'individuazione degli inattesi tramite clustering e in particolare tramite l'algoritmo DBSCAN. L'idea è di individuare le anomalie tramite un algoritmo di clustering basato sulla densità. Importi anomali o fuori periodo verrebbero visualizzati come distanti e quindi clusterizzati separatamente rispetto agli atteggiamenti standard.
2. Raggruppamenti comportamentali per i soggetti giuridici in funzione del Ramo di Attività Economica e Settore di Attività Economica. Questa suddivisione si basa sul concetto che aziende simili tendono ad assumere atteggiamenti simili in ambito economico.
3. Estrazione dei trend. Migliorare le performance degli algoritmi rimuovendo le anomalie dovute alla stagionalità.
4. Introdurre algoritmi visivi per la ricerca delle transazioni sospette

Capitolo 9

Ringraziamenti

Sono stato in dubbio se scrivere o meno questo paragrafo fino a dieci minuti prima della stampa. Quello che mi ha convinto è stato un messaggio da parte della mia ragazza, Laura, che mi ha chiesto “ti posso aiutare in qualche modo?”. Mi sono reso conto che non sarei mai giunto a questo punto senza di lei, dunque il minimo che possa fare è ringraziarla qui, per iscritto, in modo che rimanga stampato su queste pagine. È lei che mi ha spronato, che mi ha dato le energie per proseguire testardamente e per portare a degno compimento un capitolo importante della mia vita, così da poterne iniziare uno nuovo in sua compagnia.

Ringrazio anche non tanto l’azienda, che è un’entità astratta, quanto le persone concrete che la compongono, e in particolare coloro che mi hanno permesso di fare lo stage e dunque la tesi e che mi hanno supportato (e sopportato) in questo periodo, ovvero Enrico e Massimo. Grazie, davvero.

E ringrazio Emanuele, che mi ha consigliato e istruito, Simone, che mi ha aiutato a distrarmi dalle preoccupazioni e Marco, senza il quale non sarei mai andato a Pisa. Ringrazio tutti coloro che mi hanno visto dormire poco, che mi hanno visto stressato e nervoso. Ringrazio coloro che mi hanno offerto un caffè.

Ma è la mia famiglia che ringrazio, più di chiunque altro. La mia famiglia che mi è stata accanto e che mi ha guidato. Mio padre, mia madre, le mie sorelline Susanna e Veronica, mio fratello Lorenzo, mia nonna e mia zia.

E i miei quattro gatti, che mi hanno sostenuto con un sacco di fusa.

Appendice A

L'Archivio Unico Informatico

L'Archivio Unico può contenere tre tipologie di registrazioni: le operazioni, i rapporti e i legami. I rapporti e i legami riguardano le relazioni che intercorrono tra il cliente e l'intermediario finanziario o tra il cliente e gli altri clienti, e non verranno trattate in questa sede. Se si desiderassero informazioni aggiuntive sarà possibile approfondire consultando il *Provvedimento recante disposizioni attuative per la tenuta dell'archivio unico informatico e per le modalità semplificate di registrazione di cui all'art. 37, commi 7 e 8, del decreto legislativo 21 novembre 2007, n. 231* emanato dalla Banca d'Italia il 23 dicembre 2009. Nel caso di un'operazione i Tipi Informazione sono quelli riportati nella tabella A.1.

La struttura del Tipo Informazione uno, sempre nel caso di operazione, conterrà le voci della tabella A.2.

Riguardo le anagrafiche lo schema è uguale per tutti i Tipi Informazione. Le uniche differenze sono che i Tipi Informazione 4 e 6 non necessitano della Classificazione Soggetto (campi D21, D22, D23) e dei Dati sul rapporto movimentato (D31, D32), il Tipo Informazione cinque a seguito della voce riguardante il domicilio (D16) non richiede altro che il rapporto (D31) mentre solamente i Tipi Informazione 2 e 3 utilizzano completamente la tabella A.3.

Tipo Informazione	Descrizione
1	Dati caratteristici dell'operazione
2	Soggetto che ha eseguito l'operazione in proprio
3	Soggetto per conto del quale l'operazione è stata eseguita
4	Soggetto che ha eseguito l'operazione per conto terzi
5	Soggetto controparte
6	Soggetto titolare di un'operazione di cui all'art. 9, comma 1, lett. a), 2, 3, e 4 del provvedimento

Tabella A.1: I vari Tipi Informazione

Codice	Descrizione	Pic	Lng	Obb
A01	Intermediario Che Esegue La Registrazione			
A01.A	Tipo Intermediario	X	2	O
A01.B	Codice Intermediario	X	11	O
A02	Tipo Informazione	X	1	O
A03	Identificativo Registrazione	X	20	O
	Dipendenza Interessata Dall'operazione			
A11	Codice Interno Della Dipendenza	X	6	O
A12	Comune Della Dipendenza Interessata			
A12.A	Codice Cab Comune	X	6	O
A12.B	Descrizione In Chiaro	X	30	O
A12.C	Sigla Provincia	X	2	O
	Caratteristiche Dell'operazione			
A21	Data Dell'operazione	X	8	O
A22	Flag Frazionata/Multipla	X	1	O
A23	Flag Contanti	X	1	O
A24	Codice Causale Analitica	X	4	O
A25	Codice Voce	X	2	C
A31	Codice Intermediario			
A31.A	Tipo Intermediario	X	2	C
A31.B	Codice (Secondo Il Tipo)	X	11	C
A32	Paese Estero Dell'intermediario	X	3	C
A33	Comune Dell' Intermediario Della Controparte			
A33.A	Codice Cab	X	6	C
A33.B	Descrizione In Chiaro	X	30	C
A33.C	Sigla Provincia	X	2	C
A34	Descrizione Intermediario Della Controparte	X	50	C
A41	Rapporto	X	25	C
A42	Tipo Rapporto	X	1	C
A43	Tipo Legame	X	1	C
A51	Data Di Registrazione	X	8	O
A52	Tipo Di Registrazione	X	2	O
A53	Connessione Tra Registrazioni	X	12	C
A54	Stato Della Registrazione			
A54.A	Codice Stato	X	1	O
A54.B	Data Della Rettifica	X	8	C
A54.C	Identificativo Registrazione Modificata	X	20	C
B11	Euro O Divise Non Uem	X	1	O
B12	Codice Divisa	X	3	O
B13	Segno Monetario	X	1	O
B14	Importo Totale	9	15	O
B15	Importo "Di Cui Contante"	9	15	C
	A Disposizione	X	69	F

Tabella A.2: I campi del Tipo Informazione 1

Codice	Descrizione	Pic	Lng	Obb
	Identificativo Informazione			
A01	Intermediario Che Esegue La Registrazione			
A01.A	Tipo Intermediario	X	2	O
A01.B	Codice Intermediario	X	11	O
A02	Tipo Informazione	X	1	O
A03	Identificativo Registrazione	X	20	O
C11	Ndg/Connessione Con Anagrafe	X	16	C
	Complete Generalità			
D11	Cognome E Nome/Ragione Sociale	X	70	O
D13	Paese Di Residenza	X	3	O
D14	Comune Di Residenza Anagrafica			
D14.A	Codice Cab Comune	X	6	C
D14.B	Descrizione In Chiaro	X	30	C
D14.C	Sigla Provincia	X	2	C
D15	Domicilio/Sede: Indirizzo	X	35	C
D16	Domicilio/Sede: Cap Di Residenza	X	5	C
D17	Codice Fiscale	X	16	C
D18	Data Di Nascita Persone Fisiche	X	8	C
D19	Comune/Paese Di Nascita	X	30	C
	Classificazione Soggetto			
D21	Sottogruppo Di Attività Economica	X	3	O
D22	Ramo/Gruppo Di Attività Economica	X	3	O
D23	Settorizzazione Sintetica Economica	X	3	O
	Dati Sul Rapporto Movimentato			
D31	Rapporto	X	25	C
D32	Tipo Rapporto	X	1	C
	Estremi Del Documento Di Identificazione			
D41	Tipo Documento Presentato	X	2	C
D42	Numero Documento Presentato	X	15	C
D43	Data Di Rilascio	X	8	C
D44	Autorità E Località Di Rilascio	X	30	C
D45	Sesso	X	1	C
	A Disposizione	X	34	F

Tabella A.3: I campi dei Tipi Informazione 2-6

Appendice B

Gli indici di anomalia

- Indici di anomalia relativi a tutte le categorie di operazioni
 - Ripetute operazioni della stessa natura non giustificate dall'attività svolta dal cliente ed effettuate con modalità tali da denotare intenti dissimulatori
 - Ricorso a tecniche di frazionamento dell'operazione, soprattutto se volte a eludere gli obblighi di identificazione e registrazione
 - Operazioni di ingente ammontare che risultano inusuali rispetto a quelle di norma effettuate dal cliente, soprattutto se non vi sono plausibili giustificazioni economiche o finanziarie
 - Operazioni con configurazione illogica, soprattutto se risultano svantaggiose per il cliente sotto il profilo economico o finanziari
 - Operazioni effettuate frequentemente da un cliente in nome o a favore di terzi, qualora i rapporti non appaiono giustificati
 - Operazioni effettuate da terzi in nome o a favore di un cliente senza plausibili giustificazioni
 - Operazioni richieste con indicazioni palesemente inesatte o incomplete, tali da far ritenere l'intento di occultare informazioni essenziali, soprattutto se riguardanti i soggetti interessati dall'operazione
 - Operazioni con controparti insediate in aree geografiche note come centri off-shore o come zone di traffico di stupefacenti o di contrabbando di tabacchi, che non siano giustificate dall'attività economica del cliente o da altre circostanze
- Indici di anomalia relativi alle operazioni in contante e con moneta elettronica
 - Prelevamento di denaro contante per importi rilevanti, salvo che il cliente non rappresenti particolari esigenze

- Versamento di denaro contante per importi rilevanti, non giustificabile con l'attività economica del cliente
 - Ricorso al contante in sostituzione degli usuali mezzi di pagamento utilizzati dal cliente
 - Cambio di banconote con banconote di taglio diverso e/o di altre valute, soprattutto se effettuato senza transito per il conto corrente
 - Operazioni aventi a oggetto l'utilizzo di moneta elettronica che, per importo o frequenza, non risultano coerenti con l'attività svolta dal distributore o dal merchant ovvero con il normale utilizzo dello strumento da parte della clientela
- Indici di anomalia relativi alle operazioni in strumenti finanziari e alle polizze assicurative
 - Negoziazione di strumenti finanziari senza che l'operazione transiti sul conto corrente del cliente
 - Negoziazioni di strumenti finanziari aventi scarsa diffusione tra il pubblico, ripetute con elevata frequenza e/o di importo rilevante, soprattutto se concluse con controparti insediate in Paesi non comunitari ovvero non appartenenti all'OCSE
 - Ricorso a tecniche di cointestazione dei contratti aventi a oggetto strumenti finanziari o delle polizze assicurative ovvero variazioni delle intestazioni degli stessi senza plausibili giustificazioni
- Indici di anomalia relativi alle polizze assicurative vita e ai rapporti di capitalizzazione
 - Stipulazione di diverse polizze di assicurazione con pagamento dei relativi premi mediante assegni bancari che presentano molteplici girate
 - Stipulazione di polizza di assicurazione sulla vita con beneficiario il portatore della polizza
 - Nomina di più beneficiari di polizze assicurative in modo tale che l'importo da liquidare risulti frazionato in tranche, non giustificata dai rapporti tra il cliente e i beneficiari
 - Liquidazione in un arco temporale ravvicinato di prestazioni relative a molteplici polizze sottoscritte da clienti diversi e aventi come beneficiario la stessa persona
 - Rilevanti e/o contemporanee richieste di riscatto e/o di prestito relative a più polizze assicurative, soprattutto qualora comportino l'accettazione di condizioni non convenienti, ovvero frequenti operazioni di riscatto parziale relative a polizze a premio unico di rilevante importo

- In caso di pagamento di premi di rilevante importo, esercizio del diritto di revoca o del diritto di recesso di cui agli artt. 111 e 112 del d.lgs. 174/95
- Stipulazione di un contratto di capitalizzazione con consegna da parte del contraente di titoli o altri beni (v. art. 40 d.lgs. 174/95) il cui possesso non sia giustificato dalla capacità economica e dall'attività svolta dallo stesso
- Indici di anomalia relativi alle operazioni in altri prodotti e servizi
 - Presentazione di libretti di deposito bancari o postali al portatore con saldo superiore al limite di legge ancora in circolazione, senza che il cliente fornisca adeguate spiegazioni sulla tardiva presentazione degli stessi
 - Utilizzo di lettere di credito e altri sistemi di finanziamento commerciale per trasferire somme tra Paesi, senza che la relativa transazione sia giustificata dall'usuale attività economica del cliente
 - Intestazione fiduciaria di beni e/o di strumenti finanziari qualora gli stessi risultino in possesso del cliente da un breve intervallo di tempo quando ciò non appaia giustificato in relazione alla situazione patrimoniale del cliente o all'attività svolta
 - Ripetuti utilizzi di cassette di sicurezza o di servizi di custodia o frequenti depositi e ritiri di plichi sigillati, non giustificati dall'attività o dalle abitudini del cliente
 - Rilascio di deleghe a operare su cassette di sicurezza a terzi non facenti parte del nucleo familiare o non legati da rapporti di collaborazione o di altro tipo idonei a giustificare tale rilascio
 - Acquisto o vendita di rilevanti quantità di monete, metalli preziosi o altri valori, senza apparente giustificazione e/o non in linea con le condizioni economiche del cliente
 - Rapporti che presentano una movimentazione non giustificata dall'attività svolta dal cliente e che risultano caratterizzati da: versamenti frequenti di assegni o presentazione allo sconto di titoli, soprattutto se in cifra tonda, con pluralità di girate, con altri elementi ricorrenti ovvero emessi al portatore o a favore dello stesso traente; richiami dei titoli e ritorni di insoluti a volte seguiti da protesto; sostanziale pareggiamento degli addebiti e degli accrediti
- Indici di anomalia relativi al comportamento della clientela codice dell'anti-riciclaggio
 - Clienti che si rifiutano o si mostrano ingiustificatamente riluttanti a fornire le informazioni occorrenti per l'effettuazione delle operazioni, a

- dichiarare le proprie attività, a presentare documentazione contabile o di altro genere, a segnalare i rapporti intrattenuti con altri intermediari, a dare informazioni che, in circostanze normali, renderebbero il cliente stesso idoneo a effettuare operazioni bancarie, finanziarie o assicurative
- Clienti che chiedono di ristrutturare l'operazione quando la configurazione originariamente prospettata implichi forme di identificazione o registrazione oppure supplementi di istruttoria da parte dell'intermediario
 - Clienti che evitano contatti diretti con i dipendenti o i collaboratori dell'intermediario rilasciando deleghe o procure in modo frequente e ingiustificato
 - Clienti che presentano materialmente titoli o certificati per ingenti ammontari, soprattutto se al portatore, ovvero che, a seguito di operazioni di acquisto, ne richiedono la consegna materiale
 - Clienti che senza fornire plausibili giustificazioni si rivolgono a un intermediario o a un suo collaboratore lontani dalla zona di residenza o di attività, soprattutto se richiedono la domiciliazione della corrispondenza presso lo stesso
 - Clienti che effettuano operazioni di importo significativo con utilizzo di contante o strumenti al portatore quando risulti che gli stessi sono stati recentemente sottoposti ad accertamenti disposti nell'ambito di procedimenti penali o per l'applicazione di misure di prevenzione
 - Clienti in situazione di difficoltà economica che effettuano operazioni di rilevante ammontare senza fornire plausibili giustificazioni in ordine all'origine dei fondi utilizzati
 - Clienti che richiedono di effettuare operazioni con modalità inusuali, soprattutto se caratterizzate da elevata complessità, o di importo rilevante
 - Clienti, o garanti di clienti, che frequentemente e senza fornire plausibili giustificazioni chiedono la restituzione dei valori dati in garanzia previa costituzione della provvista necessaria all'acquisto di altri strumenti finanziari
 - Clienti che richiedono o intrattengono con gli intermediari rapporti con configurazione illogica

Bibliografia

- [Bel98] Franco Belli, Francesco Mazzini,
Argomenti di diritto bancario
Giappichelli Editore, 1998
- [Ber98] Peter L. Bernstein,
Against the gods: The remarkable story of risk.
John Wiley & Sons Inc, USA, 2007
- [Ber96] F. Berghella, G. Conforti, U. Fava, A. Lo Monaco,
Guida agli adempimenti antiriciclaggio
Bancaria Editrice, 1996
- [Bot11] Ezio Bottarelli,
Quaderno di Epidemiologia veterinaria,
2011
<http://www.quadernodiepidemiologia.it/epi/HomePage.html>
- [Cia08] Nicola Ciaramella,
Dispense di Sistemi Informativi Aziendali,
2008
- [Dip08] Clara Di Pillo,
Le nuove norme sull'antiriciclaggio,
2008
http://legale.guidaconsumatore.com/0068_le-nuove-norme-sullantiriciclaggio/
- [Fal09] Giovanni Falcone,
L'Antiriciclaggio: Evoluzione storica del delitto di riciclaggio di denaro sporco.
http://www.tidona.com/pubblicazioni/marzo04_4.htm,

- 2004.
- [Log09] Wikipedia contributors,
Lognormal distribution
Wikipedia, The Free Encyclopedia, 2009
http://en.wikipedia.org/wiki/Lognormal_distribution
- [1] Oasi,
Gianos®©-3D Gestione adempimenti Terza Direttiva UE,
2009
- [Paa10] Ameeth Paatil,
Running pentaho transformation stored in DB repository from Java,
2010
<http://ameethpaatil.blogspot.com/2010/11/pentaho-data-integration-java-maven.html>
- [Pow07] Daniel J. Power,
A Brief History of Decision Support Systems.
DSSResources.COM, World Wide Web,
<http://DSSResources.COM/history/dsshhistory.html>,
version 4.0, 2007
- [SAG09] Wikipedia contributors,
Semi Automatic Ground Environment
Wikipedia, The Free Encyclopedia, 2009
http://en.wikipedia.org/w/index.php?title=Semi_Automatic_Ground_Environment&oldid=330512124
- [Sco06] F. Scotti,
Il test di Student.
http://www.dti.unimi.it/fscotti/ita/md_biotec_estrazione/allegati/Student.pdf,
2006
- [Stu09] Wikipedia contributors,
Student's t-distribution
Wikipedia, The Free Encyclopedia, 2009
http://en.wikipedia.org/wiki/Student's_t-distribution

- [Vil08] Maurizio Villani,
La nuova normativa antiriciclaggio e i suoi effetti sul piano della lotta all'evasione fiscale
Fisco e Tasse, 2008
<http://www.fiscoetasse.com/approfondimenti/10042-la-nuova-normativa-antiriciclaggio-e-i-suoi-effetti-sul-piano-della.html>
- [Wil09] Williams, V., Chan, B.,
Using SAS IOM with a Java client application,
2009
- [Wri10] Wright, J., Barnes Nelson, G.,
Developing and Deploying Java Applications Around SAS: What they didn't tell you in class,
2010