

REPORT ON RISK CALCULATION MODELS IN MPS

Preparato da:

Joana Curri

Approvato da:

Mario Turla

Registro e Storia delle Revisioni				
Rev.	Data	Sezioni	Descrizione	Emiss./Appr.
0	28 .11.18	Tutte	Prima emissione	Curri/Turla

Registro e Storia delle Revisioni				
Rev.	Data	Sezioni	Descrizione	Emiss./Appr.
0	28 .11.18	Tutte	Prima emissione	Curri/Turla

Sommario

1. PREMESSA ED OBIETTIVO DEL DOCUMENTO.....	3
2. AMBITI DI APPLICAZIONE.....	3
3. METODOLOGIA ADOTTATA	4
3.1 CLUSTER.....	4
3.2 REA.....	6
3.3 PRATICHE	6
3.4 GENERAZIONE AUTOMATICA DELLA PRATICA DI SOS.....	7
3.5 GENERAZIONE MANUALE DELLA SOS.....	7
4. TEORIA BAYESIANA	8
5. CREAZIONE MODELLO	8
6. AFFIDABILITÀ DEI MODELLI DI RISCHIO APPLICATI	10
6.1 AZIENDE CHE VINCONO APPALTI CON E SENZA CIG.....	10
6.2 CONCESSIONARIE DI APPALTI	13
6.3 DIPENDENTI PUBBLICI	15
6.4 FINANZA ISLAMICA.....	17
6.5 FONDAZIONE ISLAMICA	19
6.6 HAWALLA.....	21
6.7 CARITÀ ISLAMICA	23
6.8 GENERICO	24
6.9 RICICLAGGIO	27
7. VALUTAZIONE FINALE.....	28
8. REGOLE EVIDENZE ANOMALIA.....	29
9. WORKFLOW DI LAVORAZIONE.....	39
10. GLOSSARIO DEI TERMINI UTILIZZATI.....	40

1. PREMESSA ED OBIETTIVO DEL DOCUMENTO

L'oggetto di analisi preso in considerazione è il rischio reale, in particolare quello di riciclaggio di denaro, di finanziamento al terrorismo e di corruzione.

La strada che è stata seguita per prevenire le diverse classi di rischio sul finanziamento al terrorismo e corruzione è indubbiamente quella probabilistica, per via dell'aleatorietà del rischio stesso, essendo quest'ultimo una situazione che non rispetta i principi fondamentali di tracciabilità e trasparenza.

Il processo di modellizzazione prevede la creazione di modelli, ossia l'elaborazione dei dati attraverso l'adattamento di modelli statistici, che nel nostro ambito sono caratterizzati da fenomeni probabilistici.

Ma non è da escludere un percorso di tipo deterministico, in quanto sono presenti un numero totale di variabili che portano il rischio direttamente ad alto, qualora è presente una determinata categoria nelle variabili.

Infatti, in seguito ad un'analisi dei dati, è stato ritenuto opportuno ampliare il numero di modelli probabilistici, ovvero non focalizzare più l'attenzione solo sui due macro gruppi, quali corruzione e terrorismo, ma approfondire l'analisi in più sotto ambiti, al fine di migliorarne l'affidabilità.

2. AMBITI DI APPLICAZIONE

Di seguito sono riportati gli ambiti oggetto di studio e i relativi sotto ambiti ai quali sono stati applicate le tecniche statistiche probabilistiche con fini previsivi.

1. CORRUZIONE
 - 1.1 Aziende che vincono appalti con e senza CIG
 - 1.2 Concessionarie di appalti
 - 1.3 Dipendenti Pubblici
2. TERRORISMO
 - 2.1 Finanza islamica
 - 2.2 Fondazione islamica
 - 2.3 Hawalla
 - 2.4 Carità
3. RICICLAGGIO:
 - 3.2 Società opache
 - 3.3 Bonifici per cassa
 - 3.4 Contanti
4. GENERICO

3. METODOLOGIA ADOTTATA

Ciascun rischio è stato classificato per ambito di competenza:

1. Corruzione;
2. Finanziamento al terrorismo;
3. Riciclaggio.

Per ciascun ambito sono stati individuati i principali sotto ambiti a cui viene applicato uno specifico modello probabilistico affinché siano prodotte previsioni di rischio adeguate alle singole specificità ed alle etichette di classe.

L'applicazione di un solo modello di rischio comune a tutti i sotto ambiti, come definiti sopra, non sarebbe rappresentativo in quanto un modello probabilistico "generico, teso ad identificare per esempio il rischio connesso al finanziamento del terrorismo (ambito), non consentirebbe di classificare correttamente eventi (sotto ambiti) diversi quali hawalla e la carità islamica, che hanno caratteristiche simili ma fenomenologie diverse pur in presenza dei medesimi attributi.

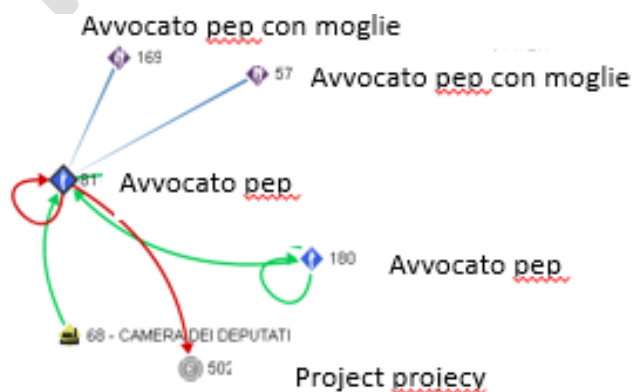
Infatti, il modello relativo al sotto ambito "carità islamica" sarà caratterizzato da rischio alto qualora taluni attributi (p.e. operazioni di importo relativamente basso) , mentre gli stessi valori assumono nel modello relativo alla "Hawalla" riporterà un rischio di livello minore.

Dunque il motivo per il quale è stato costruito un modello probabilistico distinto per ogni sotto ambito, è rilevante in quanto questo problema non viene minimizzato con un solo modello generico, ma è stata apportata una metodologia più raffinata ottimizzando l'affidabilità delle singole classificazioni.

3.1 CLUSTER

L'oggetto principale all'interno di Faraday è il **Cluster**.

1. Il cluster, è un insieme di NODI collegati tra di loro da archi orientati (ovvero con un verso) detti RELAZIONI.



2. I nodi del cluster rappresentano:
 - **soggetti clienti** dell'organizzazione, ovvero che intrattengono un rapporto con la stessa
 - **soggetti esterni** all'organizzazione. (ad esempio un soggetto non cliente che detiene una relazione economica con un cliente di una Banca.)
3. Esiste un NODO/SOGGETTO particolare detto **VERTICE** da cui derivano le RELAZIONI (dirette o indirette) con altri NODI.
4. Il vertice viene individuato dal Modulo Risk Manager o direttamente dallo User tramite la funzione Cluster on Demand.
5. Il Vertice del cluster è il soggetto considerato rischioso rispetto ad un ambito e ad un sottoambito di rischio. (Es: Corruzione / Dipendenti pubblici). Da ciò ne deriverà che un cluster ha associato un Ambito e un SottoAmbito di rischio.
6. Lo stesso soggetto vertice può esserlo per cluster di Ambiti e/o sottoambiti differenti
7. una **relazione R** è individuata da una coppia di nodi (A, B) e viene indicata come $A \Rightarrow B$ o $R(A, B)$
8. una relazione ha un **verso** cioè $R(B, A)$ è l'inverso di $R(A, B)$ ad esempio:
 - a. se $R(A, B)$ significa A è tutore di B allora $R(B, A)$ è il suo inverso B è tutelato da A
 - b. la relazione può essere **riflessiva** (cappio), ovvero $A \Rightarrow A$ (es: A preleva contante) e rappresentata da una freccia che esce e rientra nello stesso nodo.
9. una relazione è caratterizzata da:
 - a. Categoria (Statica/ Dinamica)
 - b. Tipo (cointestazione, delega, etc.)
 - c. Attributi (m) - esempio (conto di cointestazione, o conto su cui insiste la delega)
10. La **Categoria** di una relazione può essere alternativamente:
 - **statica**: ovvero derivante da ruoli o legami "definiti" tra di essi (es: co-intestatari di conto, o di una polizza assicurativa, di un mutuo, delegati, deleganti, legami familiari, cariche societarie, o per i quali vi è l'evidenza di una relazione ad esempio nell'ambito di una precedente segnalazione)
 - **dinamica**: ovvero derivanti da relazioni non predefinite, ma che si rivelano analizzando l'attività di un soggetto del cluster e soggetti terzi. Questi ultimi possono essere anche esterni all'organizzazione (e quindi non esserne clienti) e

sono identificabili più "debolmente" (ad esempio nel caso delle Banche, tramite il loro IBAN che riporta Nazione, Banca, Provincia, Conto, Nome, Cognome o Ragione Sociale).

11. Ogni cluster prevede differenti **Livelli** di Relazioni. Le RELAZIONI di livello 1 sono le relazioni che:

- collegano un nodo al vertice
- collegano tra loro due nodi che sono a loro volta collegati al vertice.

Il livello 1 del cluster è pertanto costituito da:

- il vertice
- tutti soggetti che hanno relazioni dirette con il Vertice
- tutte le relazioni tra i soggetti di livello 1

Il livello 2 del cluster è costituito da:

- soggetti che hanno relazioni dirette con soggetti di primo livello, escluso il Vertice
- tutte le relazioni tra i soggetti di livello 2

Il livello 3 del cluster non viene riportato, per non rendere troppo complessa e difficile la visione del grafo oggetto di analisi e per non fuorviare l'attenzione dell'utente dalle relazioni di livelli 1 e 2. Inoltre comporterebbe una deviazione nel calcolo della probabilità di rischio dei modelli predittivi.

3.2 REA

Le regole evidenza anomalia hanno l'obiettivo di identificare tra i vari cluster quelli che hanno tenuto un comportamento rischioso.

Le REA sono applicate ai cluster in funzione di:

- sotto ambito
- fascia di rischio

Le regole evidenza anomalia hanno un peso che può essere impostato come configurazione del sistema e che INDICA in sostanza la rischiosità del comportamento.

3.3 PRATICHE

FD gestisce l'intero ciclo di vita delle pratiche di SOS (Segnalazione Operazione Sospetta) a partire da un cluster, generato in automatico o on demand, da flusso o da fd Client - (segnalazione originaria)

una precedente segnalazione inviata (segnalazione sostituiva)

Vi sono due modalità di creazione delle pratiche:

- automatica - ovvero una pratica creata direttamente dal sistema a fronte della analisi del rischio e della esecuzione delle REA con esito positivo (ovvero la REA ha evidenziato un comportamento anomalo/rischioso)
- manuale - ovvero tramite FD client su iniziativa di un utente con profilo abilitato.

In entrambi i casi la pratica generata passa attraverso un processo di autorizzazione (workflow di approvazione) prima di essere predisposta per l'invio all'UIF.

Ciascuna pratica di Segnalazione Operazione Sospetta ha associato un peso ottenuto sommando i pesi delle REA che ne hanno determinato la creazione. Il peso della Pratica permette di dare una priorità alle pratiche da lavorare che sono presentate in ordine discendente di peso nella lista pratiche.

3.4 GENERAZIONE AUTOMATICA DELLA PRATICA DI SOS

A fronte dell'esplosione del cluster, della valutazione del rischio e della esecuzione delle REA la pratica viene creata automaticamente per i cluster che soddisfano le seguenti regole:

Fascia di Rischio = Alto/Medio Alto:

Pratica creata se:

- somma dei pesi delle REA > 0 (ovvero c'è almeno una REA)
- non vi è, per lo stesso cluster, una pratica in stato finale creata negli ultimi 6 mesi e che sia associata allo stesso insieme o super-insieme di REA

Fascia di Rischio = Medio/Medio Basso/Basso:

Pratica creata se:

- somma dei pesi delle REA > 100 (parametrico impostabile nella CFG10) (ovvero ci sono almeno due REA)
- non vi è, per lo stesso cluster, una pratica in stato finale creata negli ultimi 6 mesi e che sia associata allo stesso insieme o super-insieme di REA

3.5 GENERAZIONE MANUALE DELLA SOS

La generazione manuale avviene tramite il Client FD. Solo gli utenti con profilo abilitato possono creare una SOS a partire da un cluster o da una precedente segnalazione.

4. TEORIA BAYESIANA

Essendo il rischio stesso un evento incerto, come dice il nome stesso, si prosegue con l'utilizzo di modelli probabilistici, applicando la formula del teorema di Bayes, utilizzando dunque modelli predittivi Bayesiani, in quanto sono considerati semplici ma al tempo stesso robusti ed affidabili.

Il modello Bayesiano, comporta come ipotesi iniziale l'indipendenza degli attributi, ed è costruito utilizzando la formula di Bayes delle probabilità condizionate.

Sia $Y_i = \{1,2,3,4,5\}$, la variabile di risposta, che presenta i 5 livelli di rischio, rispettivamente {basso, medio/basso, medio, medio/alto, alto} e sia $X_n = \{X_1, X_2, \dots, X_n\}$ il set delle variabili esplicative che, nel nostro caso, assumono tutte valori discreti, al fine di rendere computazionalmente più efficace l'algoritmo predittivo scelto. Dunque applicando la formula del teorema di Bayes:

$$P(Y_i | X_n) = \frac{P(X_n | Y_i) P(Y_i)}{P(X_n)}$$

Otengo la probabilità del rischio condizionata a tutte le variabili esplicative, conoscendo le probabilità a priori delle Y_i e X_n , e tale probabilità condizionata $P(X_n | Y_i)$, ossia la probabilità delle singole variabili date la variabile di risposta, tutte queste informazioni che si ottengono tramite il set di dati etichettato completamente, ossia anche per quanto riguarda la variabile di rischio.

5. CREAZIONE MODELLO

Il modello probabilistico è stato costruito in funzione della legge di Bayes, applicando per l'appunto il classificatore "naiveBayes", il quale è caratterizzato da un utilizzo semplice e veloce, ma allo stesso tempo è ritenuto robusto ed affidabile.

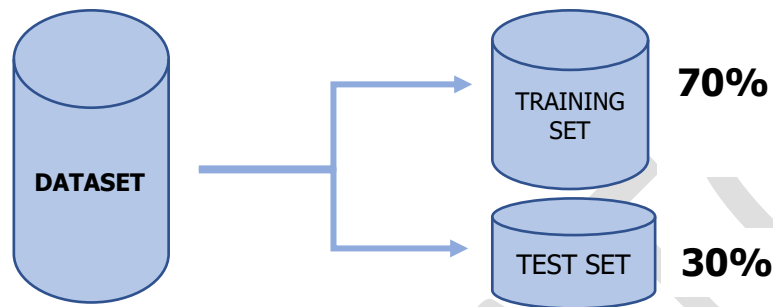
Una volta determinato l'algoritmo di previsione, ossia la regola che determina il legame tra i valori delle variabili esplicative e l'output, ovvero l'etichetta di rischio, il meccanismo è in grado di attribuire il livello di rischio a dei record nuovi che non presentano alcuna etichetta iniziale.

Proprio per questo motivo, è opportuno considerare una metodologia supervisionata per procedere alla fase di Machine Learning dell'algoritmo, tecnica questa che considera, in un set di dati, la presenza sia degli input che degli output, al fine di istruire l'algoritmo predittivo, in modo da consentirgli di risolvere dei compiti in maniera autonoma sulla base di una serie di esempi ideali, che gli vengono forniti come training.

Per poter procedere, nella maniera elencata precedentemente, è opportuno dunque avere la variabile di rischio etichettata, che in una prima fase sarà la mano umana a completare, per ogni record, i livelli di rischio opportuni, in base ai valori che presentano le variabili esplicative.

Una volta che si è in possesso di un set di dati completo dei valori della variabile di rischio, si procede con l'analisi della bontà del classificatore utilizzato, tramite la costruzione della matrice di confusione da cui si ricavano gli indici di affidabilità come l'accuratezza e l'area sotto la curva di ROC.

Il set di dati già etichettato viene suddiviso in una parte di training e una di test, rispettivamente il 70% e il 30% del set originale, al fine di addestrare il classificatore Bayesiano sul set di training mentre con il rimanente set di test vengono fatte le previsioni sul rischio, a questo punto è possibile valutare la percentuale delle classi del test set che sono state classificate correttamente tramite la costruzione della matrice di confusione, in quanto si è in possesso sia dei valori reali che di quelli predetti per quanto riguarda la variabile di etichetta del test set.



La matrice di confusione, costruita solo tramite il test set, presenta i valori di classe reali sulle righe, mentre nelle colonne si trovano i valori predetti.

In questo caso si osserva la matrice 2X2, ma è possibile estendere la variabile di risposta dicotomica ad una variabile multi-classe, nel nostro caso il numero di classi è pari a 5, dunque otterremo una matrice di confusione 5X5, e come classe positiva sarà posto il valore di rischio pari a 5, ossia il livello di rischio che corrisponde ad "ALTO".

Tramite i valori presenti all'interno della matrice è possibile ricavare determinati indici di affidabilità, tra cui:

		Predicted class	
		P	N
Actual Class	P	True Positives (TP)	False Negatives (FN)
	N	False Positives (FP)	True Negatives (TN)

- **ACCURATEZZA:** rapporto tra i valori predetti correttamente sul totale delle previsioni, misura la capacità del modello di prevedere correttamente.

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN}$$

- **ERRORE TOTALE:** indice complementare dell'accuratezza, dà informazione dell'errore di previsione commesso dal modello.

$$Error\ rate = 1 - Accuracy$$

- **SPECIFICITY:** si ottiene rapportando il numero dei veri negativi sul totale delle unità previste negative, misura quante unità previste sono realmente negative sul totale dei valori osservati negativi.

$$Specificity = \frac{TN}{TN + FN}$$

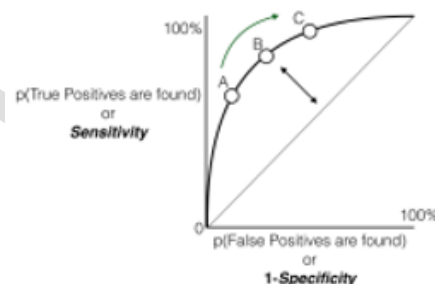
- **PRECISION:** si ottiene rapportando il numero dei veri positivi sul totale delle unità previste positive, misura la percentuale di classificazioni positive che sono corrette.

$$Precision = \frac{TP}{TP + FP}$$

- **SENSITIVITY (o RECALL):** si ottiene dal rapporto dei veri positivi e il totale delle unità previste positive, dà informazioni su quanti dati sono stimati correttamente positivi sul totale delle osservazioni positive.

$$Sensitivity = \frac{TP}{TP + FN}$$

La curva ROC viene costruita calcolando la proporzione di veri positivi (la sensibilità) e la proporzione di falsi positivi (1-specificità) per tutti i possibili valori del test. In seguito si congiungono i punti che mettono in rapporto la proporzione di veri positivi e falsi positivi. Mentre l'area sotto tale curva è una misura di quanto bene un classificatore può distinguere tra le classi del rischio, questa può assumere valori compresi tra 0.5 e 1, tanto maggiore è l'area sotto la curva tanto maggiore è il potere discriminante del test.



Qualora, dagli indici di bontà appena discussi, si ottengono risultati discreti oppure ottimali, il classificatore utilizzato risulta essere adatto a fare previsioni future su records completamente nuovi che non presentano l'etichetta di rischio.

6. AFFIDABILITÀ DEI MODELLI DI RISCHIO APPLICATI

Di seguito vengono esposti i test di affidabilità sui modelli di rischio per ciascun sotto ambito adottato.

6.1 AZIENDE CHE VINCONO APPALTI CON E SENZA CIG

Nella realizzazione delle opere pubbliche e per l'acquisizione di beni o servizi da parte della Pubblica Amministrazione è necessario indire delle gare d'appalto che permettano una gestione pubblica del territorio improntata sul rispetto della trasparenza, della concorrenza e della meritocrazia tra i vari operatori economici partecipanti. In merito a questo tema è stato ritenuto opportuno creare un opportuno modello probabilistico che attribuisce un grado di rischio alle aziende che vincono appalti con CIG, il codice identificativo gara.

Il dataset preso in considerazione presenta un numero di attributi pari a 47 e con 3133 osservazioni, mentre per rendere computazionalmente meno operoso l'algoritmo predittivo, i valori sono stati discretizzati, seguendo range differenti per ciascun sotto ambito.

I dati sono presentati in maniera strutturata, dunque nella prima riga sono presenti gli attributi mentre nella prima colonna sono riportati sia i singoli clienti e soggetti dei cluster, ma anche i singoli cluster sono stati inseriti con valori discreti ricavati facendo la media totale dei propri componenti.

Gli attributi presenti in totale sono 47, i quali si differenziano in:

- attributi anagrafici

- PAESE DI NASCITA
- FAMIGLIA CALABRESE
- PROVINCIA DI RESIDENZA
- PROVINCIA RISCHIOSA

- CITTADINANZA
- ESPOSIZIONE POLITICA
- ATECO

- attributi di rischio

- IMPORTO DARE BONIFICI ESTERI
- FREQUENZA DARE BONIFICI ESTERI
- IMPORTO AVERE BONIFICI ESTERI
- FREQUENZA AVERE BONIFICI ESTERI
- IMPORTO DARE BONIFICI ITALIA
- FREQUENZA DARE BONIFICI ITALIA
- IMPORTO AVERE BONIFICI ITALIA
- FREQUENZA AVERE BONIFICI ITALIA
- IMPORTO DARE BO. EST. FISCALITÀ AGEVOLATA
- FREQUENZA DARE BO. EST. FISCALITÀ AGEVOLATA
- IMPORTO AVERE BO. EST. FISCALITÀ AGEVOLATA
- FREQUENZA AVERE BO. EST. FISCALITÀ AGEVOLATA
- IMPORTO BONIFICI DARE CON CAUSALE RISCHIOSA
- FREQUENZA BONIFICI DARE CON CAUSALE RISCHIOSA
- IMPORTO BONIFICI AVERE CON CAUSALE RISCHIOSA
- FREQUENZA BONIFICI AVERE CON CAUSALE RISCHIOSA
- IMPORTO PRELIEVO CON CARTE CREDITO ALL'ESTERO
- FREQUENZA PRELIEVO CON CARTE CREDITO ALL'ESTERO
- IMPORTO PRELIEVO CON CARTE DEBITO IN ITALIA

- FREQUENZA PRELIEVO CON CARTE DEBITO IN ITALIA
- IMPORTO PRELIEVO CON CARTE CREDITO IN ITALIA
- FREQUENZA PRELIEVO CON CARTE CREDITO IN ITALIA
- IMPORTO SPESE CON CARTE CREDITO ALL'ESTERO
- FREQUENZA SPESE CON CARTE CREDITO ALL'ESTERO
- IMPORTO SPESE CON CARTE DEBITO IN ITALIA
- FREQUENZA SPESE CON CARTE DEBITO IN ITALIA
- IMPORTO SPESE CON CARTE CREDITO IN ITALIA
- FREQUENZA SPESE CON CARTE CREDITO IN ITALIA
- IMPORTO PRELIEVI DA CONTO CORRENTE
- FREQUENZA PRELIEVI DA CONTO CORRENTE
- IMPORTO VERSAMENTI SU RAPPORTI CONTINUATIVI
- FREQUENZA VERSAMENTI SU RAPPORTI CONTINUATIVI
- NUMERO CARTE PREPAGATE
- NUMERO CARTE DI CREDITO
- NUMERO BANCOMAT
- NUMERO DI DELEGHE PER OPERARE
- NUMERO DI CONTI
- NUMERO BANCONOTE 200 RICHIESTE E VERSATE
- NUMERO BANCONOTE 500 RICHIESTE E VERSATE
- NUMERO CASSETTE SICUREZZA

La matrice di confusione per quanto riguarda il modello sulle aziende che vincono appalti con CIG, riporta un valore di accuratezza buono, pari a 0.5767, che indica una percentuale di istanze che sono state classificate correttamente dal modello. La matrice di confusione produce inoltre come output il coefficiente statistico che rappresenta il grado di accuratezza

e affidabilità in una classificazione statistica, ossia la Kappa di Cohen. Esistono diversi "gradi di concordanza", in base ai quali possiamo definire se questo indice è ritenuto buono o meno, e il valore ottenuto dal modello probabilistico è pari a 0.3159, considerato discretamente scarsa.

Confusion Matrix and Statistics

Prediction \ True	1	2	3	4	5
1	60	63	0	0	0
2	49	81	1	0	0
3	1	3	20	3	0
4	1	2	0	9	0
5	2	1	1	0	3

Overall Statistics

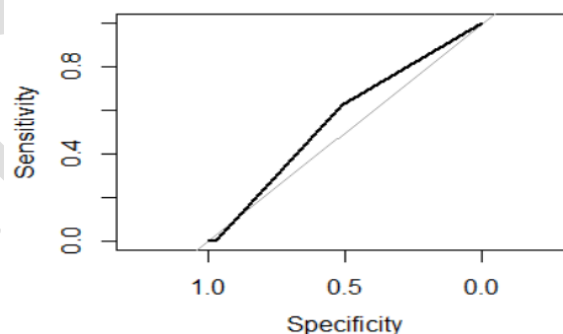
Accuracy : 0.5767
 95% CI : (0.5186, 0.6332)
 No Information Rate : 0.5
 P-Value [Acc > NIR] : 0.004631

Kappa : 0.3159
 McNemar's Test P-Value : NA

Statistics by Class:

	Class: 1	Class: 2	Class: 3	Class: 4	Class: 5
Sensitivity	0.5310	0.5400	0.90909	0.7500	1.00000
Specificity	0.6631	0.6667	0.97482	0.9896	0.98653
Pos Pred Value	0.4878	0.6183	0.74074	0.7500	0.42857
Neg Pred Value	0.7006	0.5917	0.99267	0.9896	1.00000
Prevalence	0.3767	0.5000	0.07333	0.0400	0.01000
Detection Rate	0.2000	0.2700	0.06667	0.0300	0.01000
Detection Prevalence	0.4100	0.4367	0.09000	0.0400	0.02333
Balanced Accuracy	0.5970	0.6033	0.94196	0.8698	0.99327

La curva di Roc ottenuta è la seguente, con un'area sottostante pari a 0.8405, che indica una buona performance del modello creato.



Il modello ottenuto può essere ritenuto accettabile per fare previsioni future, dal momento che più della metà delle volte fa delle previsioni corrette. Tuttavia questo col tempo andrà a migliorare la propria performance con l'aumento di nuove osservazioni ritenute corrette e modificando quelle già utilizzate che portavano il classificatore a fare previsioni non corrette.

6.2 CONCESSIONARIE DI APPALTI

Il concessionario di appalti è colui che consente di affidare ad un terzo il compito di attivare e gestire un servizio pubblico sotto il suo controllo, in cambio di una remunerazione.

Gli attributi presenti in totale sono 47, i quali si differenziano in:

- attributi anagrafici

- PAESE DI NASCITA
- FAMIGLIA CALABRESE
- PROVINCIA DI RESIDENZA
- PROVINCIA RISCHIOSA

- CITTADINANZA
- ESPOSIZIONE POLITICA
- ATECO

- attributi di rischio

- IMPORTO DARE BONIFICI ESTERI
- FREQUENZA DARE BONIFICI ESTERI
- IMPORTO AVERE BONIFICI ESTERI
- FREQUENZA AVERE BONIFICI ESTERI
- IMPORTO DARE BONIFICI ITALIA
- FREQUENZA DARE BONIFICI ITALIA
- IMPORTO AVERE BONIFICI ITALIA
- FREQUENZA AVERE BONIFICI ITALIA
- IMPORTO DARE BO. EST. FISCALITÀ AGEVOLATA
- FREQUENZA DARE BO. EST. FISCALITÀ AGEVOLATA
- IMPORTO AVERE BO. EST. FISCALITÀ AGEVOLATA
- FREQUENZA AVERE BO. EST. FISCALITÀ AGEVOLATA
- IMPORTO BONIFICI DARE CON CAUSALE RISCHIOSA
- FREQUENZA BONIFICI DARE CON CAUSALE RISCHIOSA
- IMPORTO BONIFICI AVERE CON CAUSALE RISCHIOSA
- FREQUENZA BONIFICI AVERE CON CAUSALE RISCHIOSA
- IMPORTO PRELIEVO CON CARTE CREDITO ALL'ESTERO
- FREQUENZA PRELIEVO CON CARTE CREDITO ALL'ESTERO
- IMPORTO PRELIEVO CON CARTE DEBITO IN ITALIA
- FREQUENZA PRELIEVO CON CARTE DEBITO IN ITALIA

- IMPORTO PRELIEVO CON CARTE CREDITO IN ITALIA
- FREQUENZA PRELIEVO CON CARTE CREDITO IN ITALIA
- IMPORTO SPESE CON CARTE CREDITO ALL'ESTERO
- FREQUENZA SPESE CON CARTE CREDITO ALL'ESTERO
- IMPORTO SPESE CON CARTE DEBITO IN ITALIA
- FREQUENZA SPESE CON CARTE DEBITO IN ITALIA
- IMPORTO SPESE CON CARTE CREDITO IN ITALIA
- FREQUENZA SPESE CON CARTE CREDITO IN ITALIA
- IMPORTO PRELIEVI DA CONTO CORRENTE
- FREQUENZA PRELIEVI DA CONTO CORRENTE
- IMPORTO VERSAMENTI SU RAPPORTI CONTINUATIVI
- FREQUENZA VERSAMENTI SU RAPPORTI CONTINUATIVI
- NUMERO CARTE PREPAGATE
- NUMERO CARTE DI CREDITO
- NUMERO BANCOMAT
- NUMERO DI DELEGHE PER OPERARE
- NUMERO DI CONTI
- NUMERO BANCONOTE 200 RICHIESTE E VERSATE
- NUMERO BANCONOTE 500 RICHIESTE E VERSATE
- NUMERO CASSETTE SICUREZZA

La matrice di confusione per quanto riguarda il modello sui concessionari di appalti, riporta un valore di accuratezza buono, pari a 0.5767, che indica una percentuale di istanze che sono state classificate correttamente dal modello.

La matrice di confusione produce inoltre come output il coefficiente statistico che rappresenta il grado di accuratezza e affidabilità in una classificazione statistica, ossia la Kappa di Cohen. Esistono diversi "gradi di concordanza", in base ai quali possiamo definire se questo indice è ritenuto buono o meno, e il valore ottenuto dal modello probabilistico è pari a 0.3159, considerato discretamente scarso.

Confusion Matrix and Statistics

Prediction	True				
	1	2	3	4	5
1	60	63	0	0	0
2	49	81	1	0	0
3	1	3	20	3	0
4	1	2	0	9	0
5	2	1	1	0	3

Overall Statistics

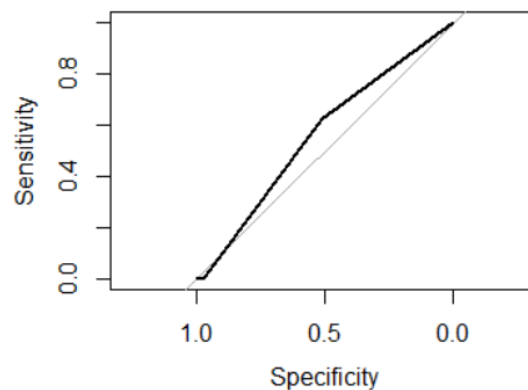
Accuracy : 0.5767
 95% CI : (0.5186, 0.6332)
 No Information Rate : 0.5
 P-Value [Acc > NIR] : 0.004631

Kappa : 0.3159
 McNemar's Test P-Value : NA

Statistics by Class:

	Class: 1	Class: 2	Class: 3	Class: 4	Class: 5
Sensitivity	0.5310	0.5400	0.90909	0.7500	1.00000
Specificity	0.6631	0.6667	0.97482	0.9896	0.98653
Pos Pred Value	0.4878	0.6183	0.74074	0.7500	0.42857
Neg Pred Value	0.7006	0.5917	0.99267	0.9896	1.00000
Prevalence	0.3767	0.5000	0.07333	0.0400	0.01000
Detection Rate	0.2000	0.2700	0.06667	0.0300	0.01000
Detection Prevalence	0.4100	0.4367	0.09000	0.0400	0.02333
Balanced Accuracy	0.5970	0.6033	0.94196	0.8698	0.99327

La curva di Roc ottenuta è la seguente, con un'area sottostante pari a 0.8405, che indica un a buona performance del modello creato.



6.3 DIPENDENTI PUBBLICI

Nell'analisi della corruzione, è emerso un sotto ambito rilevante come questo dei dipendenti pubblici, analizzando ben 38 attributi rischiosi, per individuare un rischio corruzione all'interno di soggetti che lavorano in enti pubblici.

- attributi anagrafici

- FAMIGLIA CALABRESE
- PROVINCIA DI RESIDENZA
- PROVINCIA RISCHIOSA

- ESPOSIZIONE POLITICA
- ATECO

- attributi di rischio

- IMPORTO DARE BONIFICI ITALIA
- FREQUENZA DARE BONIFICI ITALIA
- IMPORTO AVERE BONIFICI ITALIA
- FREQUENZA AVERE BONIFICI ITALIA
- IMPORTO DARE BO. EST. FISCALITÀ AGEVOLATA
- FREQUENZA DARE BO. EST. FISCALITÀ AGEVOLATA
- IMPORTO AVERE BO. EST. FISCALITÀ AGEVOLATA
- FREQUENZA AVERE BO. EST. FISCALITÀ AGEVOLATA
- IMPORTO BONIFICI DARE CON CAUSALE RISCHIOSA
- FREQUENZA BONIFICI DARE CON CAUSALE RISCHIOSA
- IMPORTO BONIFICI AVERE CON CAUSALE RISCHIOSA
- FREQUENZA BONIFICI AVERE CON CAUSALE RISCHIOSA
- IMPORTO PRELIEVO CON CARTE CREDITO ALL'ESTERO
- FREQUENZA PRELIEVO CON CARTE CREDITO ALL'ESTERO
- IMPORTO SPESE CON CARTE DEBITO IN ITALIA
- IMPORTO SPESE CON CARTE CREDITO IN ITALIA

- IMPORTO SPESE CON CARTE CREDITO SU MERCHANT A RISCHIO ITALIA
- IMPORTO SPESE CON CARTE CREDITO SU MERCHANT A RISCHIO ESTERO
- IMPORTO SPESE CON CARTE CREDITO SU MERCHANT A RISCHIO VIRTUALE
- IMPORTO SPESE CON CARTA DI CREDITO IN BENI DI LUSO IN ITALIA
- IMPORTO SPESE CON CARTA DI CREDITO IN BENI DI LUSO ALL'ESTERO
- IMPORTO SPESE CON CARTA DI CREDITO IN BENI DI LUSO VIRTUALE
- IMPORTO PRELIEVI DA CONTO CORRENTE
- FREQUENZA PRELIEVI DA CONTO CORRENTE
- IMPORTO VERSAMENTI SU RAPPORTI CONTINUATIVI
- NUMERO CARTE PREPAGATE
- NUMERO CARTE DI CREDITO
- NUMERO BANCOMAT
- NUMERO DI DELEGHE PER OPERARE
- NUMERO DI CONTI
- NUMERO BANCONOTE 200 RICHIESTE E VERSATE
- NUMERO BANCONOTE 500 RICHIESTE E VERSATE
- NUMERO CASSETTE SICUREZZA

La matrice di confusione per quanto riguarda il modello sui dipendenti pubblici, riporta un valore di accuratezza sufficientemente buono, pari a 0.6426, che indica che poco più della metà dei soggetti sono stati classificati correttamente dal classificatore, e un valore della kappa di Kohen pari al 55,48%.

Confusion Matrix and Statistics

Prediction	True				
	1	2	3	4	5
1	38	28	5	0	0
2	7	27	0	0	0
3	0	0	35	3	1
4	0	0	7	27	17
5	0	0	3	13	24

Overall Statistics

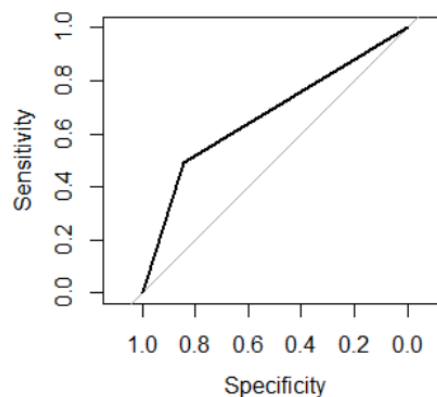
Accuracy : 0.6426
 95% CI : (0.5777, 0.7038)
 No Information Rate : 0.234
 P-Value [Acc > NIR] : < 2.2e-16

Kappa : 0.5548
 McNemar's Test P-Value : NA

Statistics by Class:

	Class: 1	Class: 2	Class: 3	Class: 4	Class: 5
Sensitivity	0.8444	0.4909	0.7000	0.6279	0.5714
Specificity	0.8263	0.9611	0.9784	0.8750	0.9171
Pos Pred Value	0.5352	0.7941	0.8974	0.5294	0.6000
Neg Pred Value	0.9573	0.8607	0.9235	0.9130	0.9077
Prevalence	0.1915	0.2340	0.2128	0.1830	0.1787
Detection Rate	0.1617	0.1149	0.1489	0.1149	0.1021
Detection Prevalence	0.3021	0.1447	0.1660	0.2170	0.1702
Balanced Accuracy	0.8354	0.7260	0.8392	0.7515	0.7443

La curva di Roc ottenuta è la seguente, con un'area sottostante pari a 0.8965, che indica una buona performance del modello creato.



6.4 FINANZA ISLAMICA

La finanza islamica è un sotto ambito del terrorismo, che cura l'aspetto finanziario strettamente connesso ai precetti religiosi, infatti negli ultimi anni ha subito un notevole sviluppo grazie alla graduale apertura, nei paesi di cultura islamica, agli investimenti esteri e, contemporaneamente, alla progressiva liberalizzazione dei sistemi bancari.

Gli attributi presenti in totale sono 32, i quali si differenziano in:

- attributi anagrafici

- PAESE DI NASCITA
- COMUNE DI NASCITA
- PROVINCIA DI RESIDENZA
- PROVINCIA RISCHIOSA

- CITTADINANZA
- ESPOSIZIONE POLITICA
- SAE
- ATECO

- attributi di rischio

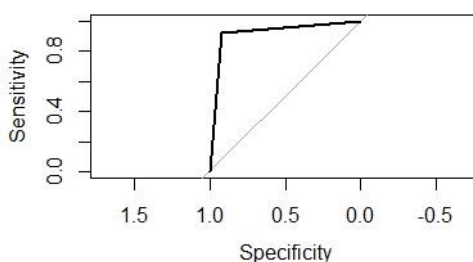
- NUMERO ASSEGNI LIBERI RICHIESTI
- IMPORTO DARE BONIFICI ESTERI
- FREQUENZA DARE BONIFICI ESTERI
- IMPORTO AVERE BONIFICI ESTERI
- FREQUENZA AVERE BONIFICI ESTERI
- IMPORTO DARE BONIFICI ITALIA
- FREQUENZA DARE BONIFICI ITALIA
- IMPORTO AVERE BONIFICI ITALIA
- FREQUENZA AVERE BONIFICI ITALIA
- IMPORTO BONIFICI DARE CON CAUSALE RISCHIOSA
- FREQUENZA BONIFICI DARE CON CAUSALE RISCHIOSA
- IMPORTO BONIFICI AVERE CON CAUSALE RISCHIOSA
- FREQUENZA BONIFICI AVERE CON CAUSALE RISCHIOSA

- IMPORTO DARE BONIFICI ITALIA A ONLUS ISLAMICHE
- FREQUENZA DARE BONIFICI ITALIA A ONLUS ISLAMICHE
- IMPORTO DARE BONIFICI ESTERO A ONLUS ISLAMICHE
- FREQUENZA DARE BONIFICI ESTERO A ONLUS ISLAMICHE
- NUMERO CARTE PREPAGATE
- NUMERO CARTE DI CREDITO
- NUMERO BANCOMAT
- NUMERO DI DELEGHE PER OPERARE
- NUMERO DI CONTI
- NUMERO SEGNALAZIONI A UIF

La matrice di confusione per quanto riguarda la finanza islamica, riporta un valore di accuratezza buono, pari a 0.7746, che indica la percentuale di istanze che sono state classificate correttamente dal modello.

La kappa di Cohen ottenuta dal modello probabilistico è pari a 0.6963, considerata dunque buona.

La curva di ROC ottenuta è la seguente, con un'area sottostante pari a 0.9437, che indica una buona performance del modello creato.



Confusion Matrix and Statistics

Prediction	Reference				
	1	2	3	4	5
1	25	1	0	0	0
2	2	11	3	1	0
3	0	0	3	0	0
4	0	0	2	14	0
5	0	0	0	7	2

Overall Statistics

Accuracy : 0.7746
 95% CI : (0.66, 0.8654)
 No Information Rate : 0.3803
 P-Value [Acc > NIR] : 1.474e-11

Kappa : 0.6963
 McNemar's Test P-Value : NA

Statistics by Class:

	Class: 1	Class: 2	Class: 3	Class: 4	Class: 5
Sensitivity	0.9259	0.9167	0.37500	0.6364	1.00000
Specificity	0.9773	0.8983	1.00000	0.9592	0.89855
Pos Pred Value	0.9615	0.6471	1.00000	0.8750	0.22222
Neg Pred Value	0.9556	0.9815	0.92647	0.8545	1.00000
Prevalence	0.3803	0.1690	0.11268	0.3099	0.02817
Detection Rate	0.3521	0.1549	0.04225	0.1972	0.02817
Detection Prevalence	0.3662	0.2394	0.04225	0.2254	0.12676
Balanced Accuracy	0.9516	0.9075	0.68750	0.7978	0.94928

6.5 FONDAZIONE ISLAMICA

La fondazione islamica, sotto ambito del terrorismo, ha un modello probabilistico che analizza le movimentazioni di denaro in paesi classificati a rischio terrorismo, o paesi limitrofi, e inoltre esamina le donazioni a Onlus islamiche italiane o estere.

Gli attributi presenti in totale sono 32, i quali si differenziano in:

- attributi anagrafici

- PAESE DI NASCITA
- COMUNE DI NASCITA
- PROVINCIA DI RESIDENZA
- PROVINCIA RISCHIOSA

- CITTADINANZA
- ESPOSIZIONE POLITICA
- SAE
- ATECO

- attributi di rischio

- NUMERO ASSEGNI LIBERI RICHIESTI
- IMPORTO DARE BONIFICI ESTERI
- FREQUENZA DARE BONIFICI ESTERI
- IMPORTO AVERE BONIFICI ESTERI
- FREQUENZA AVERE BONIFICI ESTERI
- IMPORTO DARE BONIFICI ITALIA
- FREQUENZA DARE BONIFICI ITALIA
- IMPORTO AVERE BONIFICI ITALIA
- FREQUENZA AVERE BONIFICI ITALIA
- IMPORTO BONIFICI DARE CON CAUSALE RISCHIOSA
- FREQUENZA BONIFICI DARE CON CAUSALE RISCHIOSA
- IMPORTO BONIFICI AVERE CON CAUSALE RISCHIOSA
- FREQUENZA BONIFICI AVERE CON CAUSALE RISCHIOSA

- IMPORTO DARE BONIFICI ITALIA A ONLUS ISLAMICHE
- FREQUENZA DARE BONIFICI ITALIA A ONLUS ISLAMICHE
- IMPORTO DARE BONIFICI ESTERO A ONLUS ISLAMICHE
- FREQUENZA DARE BONIFICI ESTERO A ONLUS ISLAMICHE
- NUMERO CARTE PREPAGATE
- NUMERO CARTE DI CREDITO
- NUMERO BANCOMAT
- NUMERO DI DELEGHE PER OPERARE
- NUMERO DI CONTI
- NUMERO SEGNALAZIONI A UIF

La matrice di confusione per quanto riguarda la fondazione islamica, riporta un valore di accuratezza buono, pari a 0.7746, che indica la percentuale di istanze che sono state classificate correttamente dal modello.

La matrice di confusione produce inoltre come output il coefficiente statistico che rappresenta il grado di accuratezza e affidabilità in una classificazione statistica, ossia la Kappa di Cohen. Esistono diversi "gradi di concordanza", in base ai quali possiamo definire se questo indice è ritenuto buono o meno, e il valore ottenuto dal modello probabilistico è pari a 0.6963, considerato dunque buono.

Confusion Matrix and Statistics

Prediction	Reference				
	1	2	3	4	5
1	25	1	0	0	0
2	2	11	3	1	0
3	0	0	3	0	0
4	0	0	2	14	0
5	0	0	0	7	2

Overall Statistics

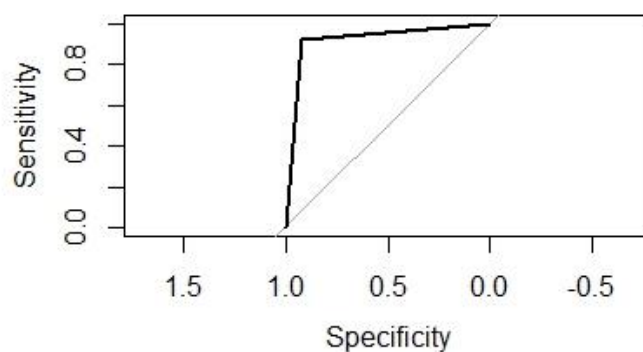
Accuracy : 0.7746
 95% CI : (0.66, 0.8654)
 No Information Rate : 0.3803
 P-Value [Acc > NIR] : 1.474e-11

Kappa : 0.6963
 McNemar's Test P-Value : NA

Statistics by Class:

	Class: 1	Class: 2	Class: 3	Class: 4	Class: 5
Sensitivity	0.9259	0.9167	0.37500	0.6364	1.00000
Specificity	0.9773	0.8983	1.00000	0.9592	0.89855
Pos Pred Value	0.9615	0.6471	1.00000	0.8750	0.22222
Neg Pred Value	0.9556	0.9815	0.92647	0.8545	1.00000
Prevalence	0.3803	0.1690	0.11268	0.3099	0.02817
Detection Rate	0.3521	0.1549	0.04225	0.1972	0.02817
Detection Prevalence	0.3662	0.2394	0.04225	0.2254	0.12676
Balanced Accuracy	0.9516	0.9075	0.68750	0.7978	0.94928

L'AUC è molto buona, con valore pari a 0.9437.



6.6 HAWALLA

L'Hawalla, sotto ambito del terrorismo, è un sistema di trasferimento di denaro, utilizzato per lo più per evitare di pagare le tasse ed eseguire transazioni senza destare sospetti. Si tratta di un vero e proprio sistema bancario parallelo o alternativo, che rischia di mettere in crisi anche quello legale, dunque l'ampiezza e la pericolosità del fenomeno non deve essere sottovalutata.

- **Attributi anagrafici:**

- | | |
|--------------------------|------------------------|
| – PAESE DI NASCITA | – CITTADINANZA |
| – COMUNE DI NASCITA | – ESPOSIZIONE POLITICA |
| – PROVINCIA DI RESIDENZA | – SAE |
| – PROVINCIA RISCHIOSA | – ATECO |

- **Attributi di rischio:**

- | | |
|--|--|
| – IMPORTO DARE BONIFICI ESTERI | – IMPORTO PRELIEVO CON CARTE CREDITO IN PAESI A RISCHIO TERRORISMO |
| – IMPORTO AVERE BONIFICI ESTERI | – NUMERO CARTE PREPAGATE |
| – IMPORTO BONIFICI DARE CON CAUSALE RISCHIOSA | – NUMERO CARTE DI CREDITO |
| – IMPORTO BONIFICI AVERE CON CAUSALE RISCHIOSA | – NUMERO BANCOMAT |
| – IMPORTO DARE BONIFICI ITALIA A ONLUS ISLAMICHE | – NUMERO DI DELEGHE PER OPERARE |
| – IMPORTO DARE BONIFICI ESTERO A ONLUS ISLAMICHE | – NUMERO DI CONTI |
| – IMPORTO DARE BONIFICI ESTERO VERSO PAESI A RISCHIO TERRORISMO | – NUMERO GARANZIE TOTALI |
| – IMPORTO AVERE BONIFICI ITALIA RISCHIOSI TERRORISMO | – NUMERO DI GARANZIE ESTERE |
| – IMPORTO PRELIEVO CON CARTE PREPAGATE IN PAESI A RISCHIO TERRORISMO | – NUMERO BANCONOTE 200 RICHIESTE E VERSATE |
| | – NUMERO BANCONOTE 500 RICHIESTE E VERSATE |
| | – NUMERO SEGNALAZIONI A UIF |

La matrice ottenuta riporta un grado di accuratezza "abbastanza buono", pari allo 0.66, che indica la percentuale di dati che sono stati classificati con istanza corretta rispetto al grado di rischio presupposto, tale valore può essere considerato accettabile, ma dovrà migliorarsi progressivamente aggiornando le informazioni corrette o meno.

Esistono diversi gradi di concordanza ottenuti dal modello probabilistico, in base ai quali possiamo definire se un modello risulta essere accettabile e sufficiente per poter essere applicato nel fare previsioni future.

Confusion Matrix and Statistics

	Reference				
Prediction	1	2	3	4	5
1	10	6	0	1	0
2	0	1	0	0	0
3	0	3	7	1	0
4	0	0	1	5	0
5	0	0	2	3	10

Overall Statistics

Accuracy : 0.66
 95% CI : (0.5123, 0.7879)
 No Information Rate : 0.2
 P-Value [Acc > NIR] : 2.173e-12

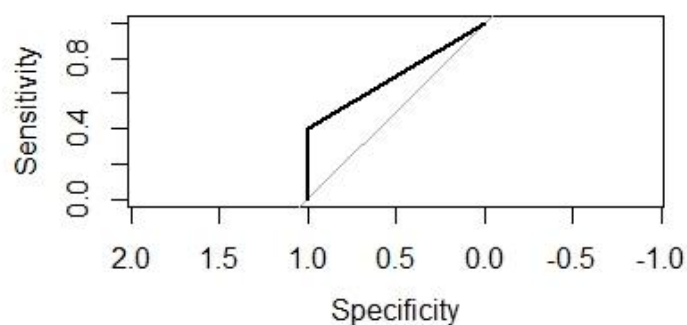
Kappa : 0.575
 McNemar's Test P-Value : NA

Statistics by Class:

	Class: 1	Class: 2	Class: 3	Class: 4	Class: 5
Sensitivity	1.0000	0.1000	0.7000	0.5000	1.0000
Specificity	0.8250	1.0000	0.9000	0.9750	0.8750
Pos Pred Value	0.5882	1.0000	0.6364	0.8333	0.6667
Neg Pred Value	1.0000	0.8163	0.9231	0.8864	1.0000
Prevalence	0.2000	0.2000	0.2000	0.2000	0.2000
Detection Rate	0.2000	0.0200	0.1400	0.1000	0.2000
Detection Prevalence	0.3400	0.0200	0.2200	0.1200	0.3000
Balanced Accuracy	0.9125	0.5500	0.8000	0.7375	0.9375

La curva di Roc, che riporta l'accuratezza del modello lungo tutto il range dei valori possibili, è la seguente, con un'area sottostante pari a 0.889.

Nel grafico si osserva l'accuratezza del modello lungo tutto il range dei valori possibili.



6.7 CARITÀ ISLAMICA

Nell'ambito del terrorismo è presente la sotto-area relativa alla carità, dove le donazioni di somme di denaro sono considerate come dei veri e propri atti rischiosi e dunque vi è richiesta una particolare attenzione a riguardo tanto da avere a disposizione di un modello probabilistico predittivo specifico per tale sotto ambito.

Nel campo note dei movimenti bancari possono essere presenti determinate parole chiavi, come donazione o carità (v. elenco nel database), che aiutano a individuare con maggiore facilità transazioni che potrebbero risultare "rischiose" in questo ambito.

Il modello probabilistico della carità è caratterizzato da importi relativamente bassi e contrariamente da frequenze abbastanza alte, particolarità questa riscontrata nelle donazioni di denaro verso centri onlus o islamici.

Gli attributi presenti in totale sono 32, i quali si differenziano in:

- attributi anagrafici

- | | |
|--------------------------|------------------------|
| – PAESE DI NASCITA | – CITTADINANZA |
| – FAMIGLIA CALABRESE | – ESPOSIZIONE POLITICA |
| – COMUNE DI NASCITA | – SAE |
| – PROVINCIA DI RESIDENZA | – ATECO |
| – PROVINCIA RISCHIOSA | |

- attributi di rischio

- | | |
|---|--|
| – NUMERO ASSEGNI LIBERI RICHIESTI | – FREQUENZA BONIFICI AVERE CON CAUSALE RISCHIOSA |
| – IMPORTO DARE BONIFICI ESTERI | – IMPORTO DARE BONIFICI ITALIA A ONLUS ISLAMICHE |
| – FREQUENZA DARE BONIFICI ESTERI | – FREQUENZA DARE BONIFICI ITALIA A ONLUS ISLAMICHE |
| – IMPORTO AVERE BONIFICI ESTERI | – IMPORTO DARE BONIFICI ESTERO A ONLUS ISLAMICHE |
| – FREQUENZA AVERE BONIFICI ESTERI | – FREQUENZA DARE BONIFICI ESTERO A ONLUS ISLAMICHE |
| – IMPORTO DARE BONIFICI ITALIA | – NUMERO CARTE PREPAGATE |
| – FREQUENZA DARE BONIFICI ITALIA | – NUMERO CARTE DI CREDITO |
| – IMPORTO AVERE BONIFICI ITALIA | – NUMERO BANCOMAT |
| – FREQUENZA AVERE BONIFICI ITALIA | – NUMERO DI DELEGHE PER OPERARE |
| – IMPORTO BONIFICI DARE CON CAUSALE RISCHIOSA | – NUMERO DI CONTI |
| – FREQUENZA BONIFICI DARE CON CAUSALE RISCHIOSA | – NUMERO SEGNALAZIONI A UIF |
| – IMPORTO BONIFICI AVERE CON CAUSALE RISCHIOSA | |

La matrice di confusione per quanto riguarda la carità, riporta un valore di accuratezza buono, pari a 0.7746, che indica che il 77,46% di istanze sono state classificate correttamente dal modello.

Il valore ottenuto dal modello probabilistico della kappa di Cohen è pari a 0.6963, considerato dunque buono.

Confusion Matrix and Statistics

	Reference				
Prediction	1	2	3	4	5
1	25	1	0	0	0
2	2	11	3	1	0
3	0	0	3	0	0
4	0	0	2	14	0
5	0	0	0	7	2

Overall Statistics

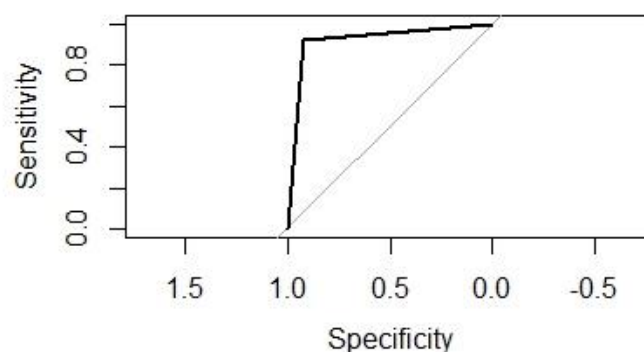
Accuracy : 0.7746
 95% CI : (0.66, 0.8654)
 No Information Rate : 0.3803
 P-Value [Acc > NIR] : 1.474e-11

Kappa : 0.6963

Statistics by Class:

	Class: 1	Class: 2	Class: 3	Class: 4	Class: 5
Sensitivity	0.9259	0.9167	0.37500	0.6364	1.00000
Specificity	0.9773	0.8983	1.00000	0.9592	0.89855
Pos Pred Value	0.9615	0.6471	1.00000	0.8750	0.22222
Neg Pred Value	0.9556	0.9815	0.92647	0.8545	1.00000
Prevalence	0.3803	0.1690	0.11268	0.3099	0.02817
Detection Rate	0.3521	0.1549	0.04225	0.1972	0.02817
Detection Prevalence	0.3662	0.2394	0.04225	0.2254	0.12676
Balanced Accuracy	0.9516	0.9075	0.68750	0.7978	0.94928

L'area sottostante alla curva di ROC è pari a 0.9437, che indica una buona performance del modello creato.



6.8 GENERICO

Nel modello generico sono presenti un numero elevato di attributi, pari a 117, di cui:

- Attributi anagrafici:

- PAESE DI NASCITA
- FAMIGLIA CALABRESE

- COMUNE DI NASCITA
- PROVINCIA DI RESIDENZA

- PROVINCIA RISCHIOSA
- CITTADINANZA
- ESPOSIZIONE POLITICA

- SAE
- ATECO

• **Attributi di rischio:**

- | | |
|--|---|
| <ul style="list-style-type: none"> - NUMERO ASSEGNI LIBERI RICHIESTI - IMPORTO DARE BONIFICI ESTERI - FREQUENZA DARE BONIFICI ESTERI - IMPORTO AVERE BONIFICI ESTERI - FREQUENZA AVERE BONIFICI ESTERI - IMPORTO DARE BONIFICI ITALIA - FREQUENZA DARE BONIFICI ITALIA - IMPORTO AVERE BONIFICI ITALIA - FREQUENZA AVERE BONIFICI ITALIA - IMPORTO DARE BO.EST. FISCALITA AGEVOLATA - FREQUENZA DARE BO.EST FISCALITA AGEVOLATA - IMPORTO AVERE BO.EST FISCALITA AGEVOLATA - FREQUENZA AVERE BO.EST FISCALITA AGEVOLATA - IMPORTO BONIFICI DARE CON CAUSALE RISCHIOSA - FREQUENZA BONIFICI DARE CON CAUSALE RISCHIOSA - IMPORTO BONIFICI AVERE CON CAUSALE RISCHIOSA - ACQUISTO VALUTE DIGITALI - IMPORTO PRELIEVO CON CARTE DEBITO ALL'ESTERO - FREQUENZA PRELIEVO CON CARTE DEBITO ALL'ESTERO - IMPORTO PRELIEVO CON CARTE PREPAGATE ALL'ESTERO - FREQUENZA PRELIEVO CON CARTE PREPAGATE ALL'ESTERO - IMPORTO PRELIEVO CON CARTE CREDITO ALL'ESTERO - FREQUENZA PRELIEVO CON CARTE CREDITO ALL'ESTERO - IMPORTO PRELIEVO CON CARTE DEBITO IN ITALIA - FREQUENZA PRELIEVO CON CARTE DEBITO IN ITALIA - IMPORTO PRELIEVO CON CARTE PREPAGATE IN ITALIA - FREQUENZA PRELIEVO CON CARTE PREPAGATE IN ITALIA - IMPORTO PRELIEVO CON CARTE CREDITO IN ITALIA - FREQUENZA PRELIEVO CON CARTE CREDITO IN ITALIA - IMPORTO SPESE CON CARTE DEBITO ALL'ESTERO - FREQUENZA SPESE CON CARTE DEBITO ALL'ESTERO - IMPORTO SPESE CON CARTE PREPAGATE ALL'ESTERO - FREQUENZA SPESE CON CARTE PREPAGATE ALL'ESTERO - IMPORTO SPESE CON CARTE CREDITO ALL'ESTERO | <ul style="list-style-type: none"> - FREQUENZA SPESE CON CARTE CREDITO ALL'ESTERO - IMPORTO SPESE CON CARTE DEBITO IN ITALIA - FREQUENZA SPESE CON CARTE DEBITO IN ITALIA - IMPORTO SPESE CON CARTE PREPAGATE IN ITALIA - FREQUENZA SPESE CON CARTE PREPAGATE IN ITALIA - IMPORTO SPESE CON CARTE CREDITO IN ITALIA - FREQUENZA SPESE CON CARTE CREDITO IN ITALIA - IMPORTO SPESE CON CARTE DEBITO SU MERCHANT a RISCHIO ITALIA - FREQUENZA SPESE CON CARTE DEBITO SU MERCHANT a RISCHIO ITALIA - IMPORTO SPESE CON CARTE PREPAGATE SU MERCHANT a RISCHIO ITALIA - FREQUENZA SPESE CON CARTE PREPAGATE SU MERCHANT a RISCHIO ITALIA - IMPORTO SPESE CON CARTE CREDITO SU MERCHANT a RISCHIO ITALIA - FREQUENZA SPESE CON CARTE CREDITO SU MERCHANT a RISCHIO ITALIA - IMPORTO SPESE CON CARTE DEBITO SU MERCHANT a RISCHIO ESTERO - FREQUENZA SPESE CON CARTE DEBITO SU MERCHANT a RISCHIO ESTERO - IMPORTO SPESE CON CARTE CREDITO SU MERCHANT a RISCHIO ESTERO - FREQUENZA SPESE CON CARTE CREDITO SU MERCHANT a RISCHIO ESTERO - IMPORTO SPESE CON CARTE PREPAGATE SU MERCHANT a RISCHIO ESTERO - FREQUENZA SPESE CON CARTE PREPAGATE SU MERCHANT a RISCHIO ESTERO - IMPORTO SPESE CON CARTE DEBITO SU MERCHANT a RISCHIO VIRTUALE - FREQUENZA SPESE CON CARTE DEBITO SU MERCHANT a RISCHIO VIRTUALE - IMPORTO SPESE CON CARTE CREDITO SU MERCHANT a RISCHIO VIRTUALE - FREQUENZA SPESE CON CARTE CREDITO SU MERCHANT a RISCHIO VIRTUALE - IMPORTO SPESE CON CARTE PREPAGATE SU MERCHANT a RISCHIO VIRTUALE - FREQUENZA SPESE CONCARTE PREPAGATE SU MERCHANT a RISCHIO VIRTUALE - IMPORTO SPESE CON CARTA DI DEBITO IN BENI DI LUSO IN ITALIA - FREQUENZA SPESE CON CARTA DI DEBITO IN BENI DI LUSO IN ITALIA |
|--|---|

- | | |
|--|--|
| <ul style="list-style-type: none"> - IMPORTO SPESE CON CARTA DI CREDITO IN BENI DI LUSO IN ITALIA - FREQUENZA SPESE CON CARTA DI CREDITO IN BENI DI LUSO IN ITALIA - IMPORTO SPESE CON CARTA PREPAGATE IN BENI DI LUSO IN ITALIA - FREQUENZA SPESE CON CARTA PREPAGATE IN BENI DI LUSO IN ITALIA - IMPORTO SPESE CON CARTA DI DEBITO IN BENI DI LUSO ALL'ESTERO - FREQUENZA SPESE CON CARTA DI DEBITO IN BENI DI LUSO ALL'ESTERO - IMPORTO SPESE CON CARTA DI CREDITO IN BENI DI LUSO ALL'ESTERO - FREQUENZA SPESE CON CARTA DI CREDITO IN BENI DI LUSO ALL'ESTERO - IMPORTO SPESE CON CARTA PREPAGATE IN BENI DI LUSO ALL'ESTERO - FREQUENZA SPESE CON CARTA PREPAGATE IN BENI DI LUSO ALL'ESTERO - IMPORTO SPESE CON CARTA DI DEBITO IN BENI DI LUSO VIRTUALE - FREQUENZA SPESE CON CARTA DI DEBITO IN BENI DI LUSO VIRTUALE - IMPORTO SPESE CON CARTA DI CREDITO IN BENI DI LUSO VIRTUALE - FREQUENZA SPESE CON CARTA DI CREDITO IN BENI DI LUSO VIRTUALE - IMPORTO SPESE CON CARTA PREPAGATE IN BENI DI LUSO VIRTUALE - FREQUENZA SPESE CON CARTA PREPAGATE IN BENI DI LUSO VIRTUALE - IMPORTO SPESE CON CARTA DI DEBITO SU MERCHANT DUAL USE - FREQUENZA SPESE CON CARTA DI DEBITO SU MERCHANT DUAL USE | <ul style="list-style-type: none"> - IMPORTO SPESE CON CARTA DI CREDITO SU MERCHANT DUAL USE - FREQUENZA SPESE CON CARTA DI CREDITO SU MERCHANT DUAL USE - IMPORTO SPESE CON CARTA PREPAGATE SU MERCHANT DUAL USE - FREQUENZA SPESE CON CARTA PREPAGATE SU MERCHANT DUAL USE - IMPORTO PRELIEVI DA CONTO CORRENTE - FREQUENZA PRELIEVI DA CONTO CORRENTE - IMPORTO VERSAMENTI SU RAPPORTI CONTINUATIVI - FREQUENZA VERSAMENTI SU RAPPORTI CONTINUATIVI - IMPORTO OPERAZIONI PER CASSA - FREQUENZA OPERAZIONI PER CASSA - SOMMA DEI CREDITI - RATING CREDITIZIO - NUMERO CARTE PREPAGATE - NUMERO CARTE DI CREDITO - NUMERO BANCOMAT - NUMERO DI DELEGHE PER OPERARE - NUMERO DI CONTI - NUMERO GARANZIE TOTALI - IMPORTO TOTALE GARANTITO - NUMERO DI GARANZIE ESTERE - IMPORTO GARANZIE ESTERE - NUMERO FIDEJUSSIONI - NUMERO BANCONOTE 200 RICHIESTE E VERSATE - NUMERO BANCONOTE 500 RICHIESTE E VERSATE - NUMERO SEGNALAZIONI A UIF - MASSIMO LIVELLO DI RISCHIO DI INOLTRO - NUMERO ALERT GENERATI IN AML - NUMERO CASSETTE SICUREZZA - FREQUENZA ACCESSO CASSETTE SICUREZZA |
|--|--|

La matrice di confusione per quanto riguarda l'ambito generico, riporta un valore di accuratezza buono, del 73.36%, che indica la percentuale di istanze che sono state classificate correttamente dal modello.

La matrice di confusione produce inoltre come output il coefficiente statistico che rappresenta il grado di accuratezza e affidabilità in una classificazione statistica, ossia la Kappa di Cohen, con valore ottenuto 0.6226.

Confusion Matrix and Statistics

	True					
Prediction	1	2	3	4	5	
1	10	28	1	0	0	
2	79	191	0	0	5	
3	10	0	75	4	2	
4	0	0	1	95	1	
5	3	6	1	1	20	

Overall Statistics

Accuracy : 0.7336
 95% CI : (0.6939, 0.7707)
 No Information Rate : 0.4221
 P-Value [Acc > NIR] : < 2.2e-16

Kappa : 0.6226
 McNemar's Test P-Value : NA

Statistics by Class:

	Class: 1	Class: 2	Class: 3	Class: 4	Class: 5
Sensitivity	0.09804	0.8489	0.9615	0.9500	0.71429
Specificity	0.93271	0.7273	0.9648	0.9954	0.97822
Pos Pred Value	0.25641	0.6945	0.8242	0.9794	0.64516
Neg Pred Value	0.81377	0.8682	0.9932	0.9885	0.98406
Prevalence	0.19137	0.4221	0.1463	0.1876	0.05253
Detection Rate	0.01876	0.3583	0.1407	0.1782	0.03752
Detection Prevalence	0.07317	0.5159	0.1707	0.1820	0.05816
Balanced Accuracy	0.51538	0.7881	0.9632	0.9727	0.84625

I modelli probabilistici predittivi hanno ottenuto una discreta capacità di performance nell'attribuire correttamente l'etichetta di classe al rischio, come si evince dai valori ottenuti degli indici di bontà, con una percentuale di classificazione corretta circa e superiore del 50%. I modelli sono adesso in grado di fare previsioni future nei vari sotto ambiti con una percentuale di classificazione errata discreta.

6.9 RICICLAGGIO

Per quanto riguarda i tre sotto ambiti del riciclaggio:

- bonifici per cassa
- società opache
- contanti

è stato applicato momentaneamente il modello probabilistico predittivo Bayesiano GENERICO, con ben 117 attributi, tra cui anagrafici e di rischio.

I risultati di performance, dunque si possono osservare nella matrice di confusione del modello precedente, tra cui l'accuratezza, la kappa di Cohen, la curva di ROC e l'area sottostante (AUC).

Nel passo successivo, questi tre sotto ambiti del riciclaggio, avranno ciascuno un modello predittivo a parte, che caratterizza le singole proprietà del sotto ambito.

7. VALUTAZIONE FINALE

Dalle analisi precedenti, si evince chiaramente che in un ambito di classificazione è importante valutare l'affidabilità del modello per fini predittivi, e ciò viene definito mediante il calcolo dell'accuratezza del classificatore attraverso vari metodi.

In questo campo di studio è stato trattato in particolar modo la divisione del set di dati in una parte più ampia di allenamento e la rimanente oggetto di valutazione.

È fondamentale chiarire alcuni concetti presupposti di base, per capire la metodologia utilizzata per considerare affidabile un modello in base ad alcuni singoli indici di bontà.

In un contesto dove la variabile di target è il livello di rischio dei singoli soggetti e/o degli interi cluster, è stato usato un approccio dove la variabile "RISCHIO" è stata costruita, ossia etichettata con valori da 1 a 5, in ordine di grado di rischio da basso ad alto, utilizzando una metodologia induttiva.

L'utilizzo dell'induzione è dovuto al fatto che operando in un contesto di incertezza non si è in grado di avere come informazione a priori il grado di rischio dei clienti, ma è questa un'informazione che viene ricavata in base ai valori che il record presenta in un tot di variabili, diverse per ogni sotto ambito.

L'approccio induttivo, permette di passare dal caso particolare al caso generale, questo concetto è fondamentale in quanto avendo analizzato nel dettaglio una serie di casi di soggetti rischiosi o meno, è stato possibile costruire un filo logico da adottare per etichettare i singoli soggetti con una classe di rischio basata sulle caratteristiche che presentano negli attributi esplicativi.

Dunque i modelli probabilistici costruiti per ciascun sotto ambito sono stati realizzati con un set di dati di partenza che non può essere completamente rappresentativo della realtà, in quanto ottenuto per induzione, e soprattutto in un contesto come questo di incertezza, è basilare che l'etichetta attribuita a ciascun record è basata molto anche sulla conoscenza pregressa di colui che assegna tale livello di rischio.

Conseguenza delle metodologie utilizzate, è il grado di accuratezza dei singoli modelli, che non risulta essere ottimale ma comunque è ritenuto sufficiente per rappresentare tali dati come primo punto di partenza.

È ovvio che il set di dati di allenamento dei modelli con il tempo sarà migliorato aggiornando alcune informazioni e/o addirittura etichettare nuovamente soggetti con un diverso grado di rischio rispetto a quanto prima fosse.

8. REGOLE EVIDENZE ANOMALIA

- Le regole di evidenza anomalia agiscono per:
- AMBITO
- SOTTOAMBITO

Ed hanno associato un peso di rischio personalizzabile per tipologia del soggetto vertice del cluster:

- Persona giuridica
- Persona fisica
- Persona non fisica

SottoAmbito	REA	PesoVPG	PesoVPF	PesoVPNF	DescrizioneSoglia	ValoreSoglia
Aziende che vincono appalti	Bonifici verso paesi a fiscalità agevolata	90	90	90	Importo totale bonifici	10000
Aziende che vincono appalti	Il cluster effettua bonifici Esteri	60	60	60	Importo Esteri	Bonifici 5000
Aziende che vincono appalti	Il cluster effettua bonifici per cassa	90	90	90	Bonifici per Cassa	10
Aziende che vincono appalti	Il cluster effettua bonifici rischiosi	50	50	50	Importo Rischiosi	Bonifici 5000
Aziende che vincono appalti	Il cluster effettua movimenti contanti	60	60	60	Movimenti contanti	30000
Aziende che vincono appalti	Il cluster utilizza banconote da 500 eur	80	80	80	Numero Banconote da 500 utilizzate	10
Aziende che vincono appalti	Il cluster utilizza moneta virtuale	90	90	90	Acquisto Virtuale	Moneta 10
Aziende che vincono appalti	Soggetti del cluster famiglie calabresi	80	80	80	Numero soggetti nel cluster in liste famiglie calabresi	0
Aziende che vincono appalti	Soggetti del cluster in liste crimini	80	80	80	Numero soggetti nel cluster in liste criminali	0

Aziende che vincono appalti	Soggetti del cluster in liste pep	90	90	90	Soggetti in liste pep o 10 criminali
Aziende che vincono appalti	Spese con carte all'estero	70	70	70	Importo spese e 1000 prelievi all'estero
Aziende che vincono appalti senza CIG	Bonifici verso paesi a fiscalità agevolata	80	80	80	Importo Totale 10000 bonifici > 10000
Aziende che vincono appalti senza CIG	Il cluster effettua bonifici Esteri	70	70	70	Importo bonifici 90000 Esteri
Aziende che vincono appalti senza CIG	Il cluster effettua bonifici per cassa	90	90	90	Bonifici per cassa 10
Aziende che vincono appalti senza CIG	Il cluster effettua bonifici rischiosi	60	60	60	Bonifici con Causali 10000 Rischiose
Aziende che vincono appalti senza CIG	Il cluster effettua movimenti contanti	70	70	70	Movimenti Contante 100000
Aziende che vincono appalti senza CIG	Il cluster utilizza banconote da 500 eur	80	80	80	Numero Banconote 10 da 500 utilizzate
Aziende che vincono appalti senza CIG	Il cluster utilizza moneta virtuale	90	90	90	Importo Totale 0 Compravendite
Aziende che vincono appalti senza CIG	Il cluster utilizza moneta virtuale	90	90	90	Compravendita 0 Moneta Virtuale come Bitcoin, etc.
Aziende che vincono appalti senza CIG	Soggetti del cluster da famiglie calabresi	80	80	80	Numero soggetti nel 0 cluster in liste famiglie calabresi
Aziende che vincono appalti senza CIG	Soggetti del cluster in liste criminali	80	80	80	Numero soggetti nel 0 cluster in liste criminali
Aziende che vincono appalti senza CIG	Soggetti del cluster in liste pep	90	90	90	Soggetti in liste PEP o 10 crime
Aziende che vincono appalti senza CIG	Spese con carte all'estero	50	50	50	Importo 1000 spese/prelievi all'estero

Concessionari di Appalti	Bonifici verso paesi a fiscalità agevolata	90	90	90	Importo Bonifici > 10000	Totale 10000
Concessionari di Appalti	Il cluster effettua bonifici Esteri	80	80	80	Il cluster Effettua Bonifici Esteri in Dare per importo > 15000	35000
Concessionari di Appalti	Il cluster effettua bonifici per cassa	90	90	90	Importo Bonifici per cassa	45000
Concessionari di Appalti	Il cluster effettua movimenti contanti	90	90	90	Importo Movimenti	Totale 20000
Concessionari di Appalti	Il cluster esegue Bon. rischiosi Importo > 10000	90	90	90	Importo Rischiosi	Bonifici 10000
Concessionari di Appalti	Il cluster utilizza banconote da 500 eur	80	80	80	Numero Banconote da 500 utilizzate	10
Concessionari di Appalti	Soggetti del cluster da famiglie calabresi	80	80	80	Numero soggetti nel cluster in liste famiglie calabresi	0
Concessionari di Appalti	Soggetti del cluster in liste criminali	80	80	80	Numero soggetti nel cluster in liste criminali	0
Concessionari di Appalti	Soggetti del cluster in liste pep	80	80	80	Numero Pep/PIL nel cluster	0
Concessionari di Appalti	Soggetti legati a edilizia, rifiuti, gioco...	70	70	70	Soggetti Rischiosi	0
Concessionari di Appalti	Spese con carte all'estero	70	70	70	Importo spese e prelievi all'estero	1000
Corruzione Dipendenti Pubblici	Il cluster effettua bonifici Esteri	80	80	80	Effettua esteri in dare per importo > di 15000	Bonifici per 15000
Corruzione Dipendenti Pubblici	Il cluster effettua bonifici per cassa	90	90	90	Importo Bonifici per cassa	10
Corruzione Dipendenti Pubblici	Il cluster effettua	70	70	70	Totale rischiosi	Bonifici 10000

Corruzione Dipendenti Pubblici	bonifici rischiosi	il cluster	70	70	70	Movimentazione in 20000 contante del cluster sopra la soglia "Importo Contanti"
Corruzione Dipendenti Pubblici	effettua movimenti contanti	Il cluster	80	80	80	Numero Banconote 10 da 500 utilizzate
Corruzione Dipendenti Pubblici	utilizza banconote da 500 eur	Il cluster	90	90	90	Importo Totale 0 Compravendite
Corruzione Dipendenti Pubblici	moneta virtuale	Il numero delle deleghe nel cluster è > 10	70	70	70	Numero deleghe del 10 vertice del cluster
Corruzione Dipendenti Pubblici	Soggetti del cluster da famiglie calabresi	80	80	80	80	Numero soggetti nel 0 cluster in liste famiglie calabresi
Corruzione Dipendenti Pubblici	Soggetti del cluster in liste criminali	80	80	80	80	Numero soggetti nel 0 cluster in liste criminali
Corruzione Dipendenti Pubblici	Soggetti del cluster in liste pep	90	90	90	90	Numero soggetti del 0 cluster che sono PEP o Criminali o legati alla ndrangheta
Corruzione Dipendenti Pubblici	soggetti del cluster utilizzano moneta virtuale	90	90	90	90	Importo Acquisto 10
Corruzione Dipendenti Pubblici	Soggetto del cluster pagato da aziende vincenti	60	60	60	60	Totale Pagamenti 10000 Ricevuti da azienda vincente appalti PA
Corruzione Dipendenti Pubblici	Spese con carte all'estero	60	60	60	60	Importo spese e 1000 prelievi all'estero
Bonifici cassa	per Bonifici verso paesi a fiscalità agevolata	80	80	80	80	Il cluster Effettua 5000 Bonifici Esteri in Dare a fiscalità agevolata per importo > 5000
Bonifici cassa	per Il cluster effettua bonifici Esteri	80	80	80	80	Il cluster Effettua 15000 Bonifici Esteri in Dare per importo > 15000

Bonifici cassa	per	Il cluster	90	90	90	Importo Bonifici per 7000 cassa	
Bonifici cassa	per	Il cluster	90	90	90	Importo Movimenti	Totale 20000
Bonifici cassa	per	Il cluster	90	90	90	Importo Rischiosi	Bonifici 10000
Bonifici cassa	per	Il cluster	80	80	80	Importo Movimenti	Totale 150000
Bonifici cassa	per	Il cluster	80	80	80	Numero Banconote da 500 utilizzate	10
Bonifici cassa	per	Il cluster	90	90	90	Importo Compravendite	Totale 0
Bonifici cassa	per	Soggetti cluster	80	80	80	Numero soggetti nel cluster in liste famiglie calabresi	0
Bonifici cassa	per	Soggetti cluster in liste criminali	80	80	80	Numero soggetti nel cluster in liste criminali	0
Bonifici cassa	per	Soggetti cluster in liste pep	80	80	80	Numero Pep/PIL nel cluster	0
Bonifici cassa	per	Spese carte all'estero	80	80	80	Importo spese e prelievi	2000
Contante	Bonifici verso paesi a fiscalità agevolata		90	90	90	Importo Bonifici	10000
Contante	Il cluster		90	90	90	Compravendita Moneta Virtuale	10000
Contante	Il cluster		80	80	80	Importo Bonifici	10000
Contante	Il cluster		90	90	90	Importo Bonifici per cassa	10

Contante	bonifici per cassa	Il cluster	50	50	50	Importo Bonifici	10000
Contante	effettua bonifici rischiosi	Il cluster	90	90	90	Importo Movimenti Totale	20000
Contante	effettua movimenti contanti	Il cluster ha SAE che muovono contanti	80	80	80	Importo Contanti	150000
Contante	utilizza banconote da 500 eur	Il cluster	90	90	90	Numero Banconote da 500 eur	10
Contante	Soggetti cluster famiglie calabresi	del cluster da	80	80	80	Numero soggetti nel cluster in liste famiglie calabresi	0
Contante	Soggetti cluster in liste criminali	del cluster in liste	80	80	80	Numero soggetti nel cluster in liste criminali	0
Contante	Soggetti cluster in liste pep	del cluster in liste	80	80	80	Numero Pep/PIL nel cluster	0
Contante	Spese carte all'estero	con	80	80	80	Importo spese e prelievi	2000
Controllo Livello	II Banconote da 500		80	80	80	Importo	10
Controllo Livello	II Bon.Esteri Fiscalità Agevolata		80	80	80	Importo	10000
Controllo Livello	II Bonifici Esteri		80	80	80	Importo	10000
Controllo Livello	II Bonifici per Cassa		80	80	80	Importo	10000
Controllo Livello	II Bonifici Rischiosi		80	80	80	Importo	10000
Controllo Livello	II Criminali		80	80	80	Importo	0
Controllo Livello	II Famiglie Calabresi		80	80	80	Importo	0
Controllo Livello	II Movimento Contanti		80	80	80	Importo	150000
Controllo Livello	II Numero Deleghe		80	80	80	Importo	10

Controllo Livello	II	Pep/PIL	80	80	80	Importo	0
Controllo Livello	II	SAE600 Muovono contanti	80	80	80	Importo	150000
Controllo Livello	II	Spese con Carte all'estero	80	80	80	Importo	10000
Controllo Livello	II	Valuta Virtuale	80	80	80	Importo	10000
Società Opache		Bonifici verso paesi a fiscalità agevolata	90	90	90	Importo Bonifici > 10000	Totale 10000
Società Opache		Il cluster effettua Bonifici Esteri	80	80	80	Il cluster effettua bonifici esteri per importi > 15.000 EUR	10000
Società Opache		Il cluster effettua bonifici per cassa	90	90	90	Importo Bonifici per cassa	10
Società Opache		Il cluster effettua bonifici rischiosi	70	70	70	Importo Rischiosi	Bonifici 5000
Società Opache		Il cluster effettua movimenti contanti	70	70	70	Movimentazione contante del cluster sopra la soglia	in 20000
Società Opache		Il cluster ha SAE600 che muovono contanti	80	80	80	Importo Movimenti	Totale 150000
Società Opache		Il cluster utilizza banconote da 500 eur	80	80	80	Numero Banconote da 500 utilizzate	10
Società Opache		Il cluster utilizza moneta virtuale	90	90	90	Importo Compravendite	Totale 0
Società Opache		Soggetti del cluster famiglie calabresi	80	80	80	Numero soggetti nel cluster in liste famiglie calabresi	0
Società Opache		Soggetti del cluster in liste criminali	80	80	80	Numero soggetti nel cluster in liste criminali	0

Società Opache	Soggetti del cluster in liste pep	90	90	90	Numero Soggetti PEP 0 o Criminali contenuti nel cluster > 0
Società Opache	Spese con carte all'estero	80	80	80	Importo totale spese 1000 e prelievi all'estero
Carità	Bonifici verso paesi a fiscalità agevolata	90	90	90	Importo Bonifici Totale 5000
Carità	Il cluster effettua Bonifici Esteri	70	70	70	Importo bonifici totale 5000
Carità	Il cluster effettua bonifici per cassa	90	90	90	Importo Bonifici per cassa 10
Carità	Il cluster effettua movimenti contanti	90	90	90	Importo Movimenti Totale 20000
Carità	Il cluster esegue Bon. rischiosi	90	90	90	Importo Rischiosi Bonifici 10000
Carità	Il cluster ha SAE600 che muovono contanti	80	80	80	Importo Movimenti Totale 150000
Carità	Il cluster utilizza banconote da 500 eur	80	80	80	Numero Banconote da 500 utilizzate 10
Carità	Il cluster utilizza moneta virtuale	90	90	90	Importo Compravendite Totale 0
Carità	Soggetti del cluster famiglie calabresi	80	80	80	Numero soggetti nel cluster in liste famiglie calabresi 0
Carità	Soggetti del cluster in liste crimini	80	80	80	Numero soggetti nel cluster in liste criminali 0
Carità	Soggetti del cluster in liste pep	90	90	90	Numero di criminali o PEP nel cluster 0

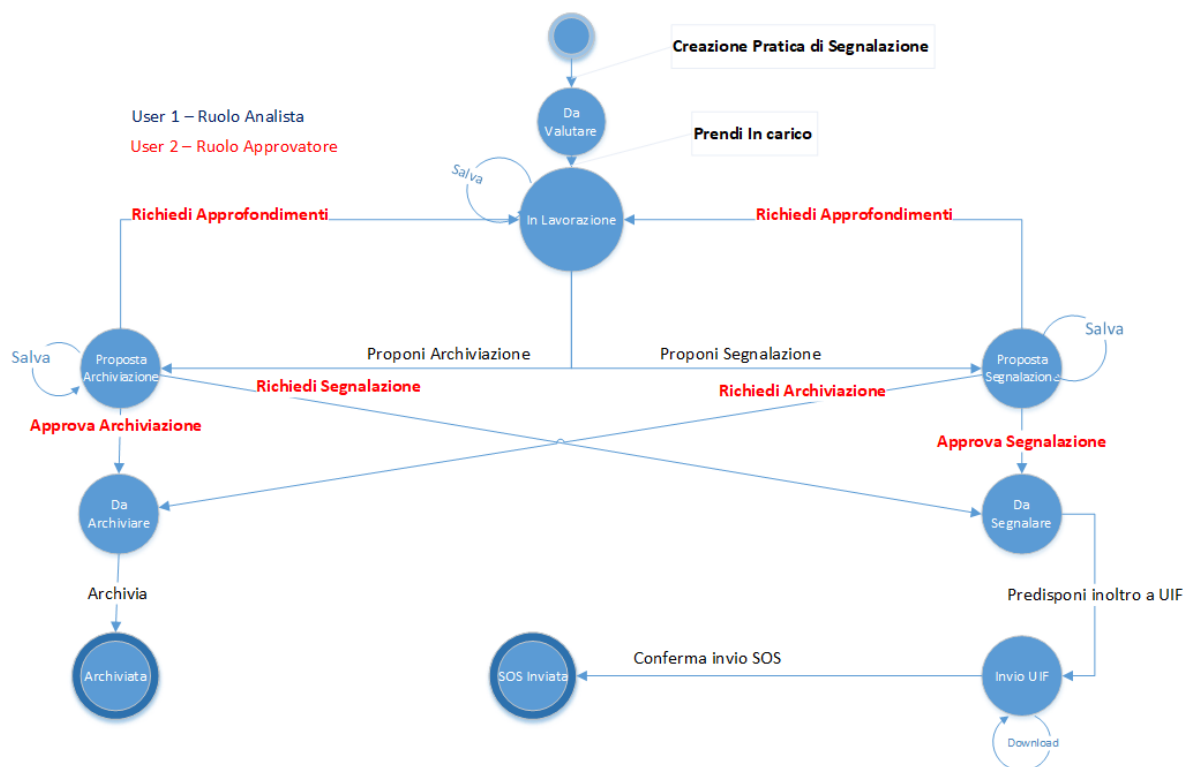
Carità	Spese con 80	80	80	Importo totale spese 1000 e prelievi
Finanza islamica	Bonifici verso paesi a fiscalità agevolata	90	90	Importo bonifici totale 5000
Finanza islamica	Il cluster 80	80	80	Importo Esteri Bonifici 5000
Finanza islamica	Il cluster 90	90	90	Importo Bonifici per 10 cassa
Finanza islamica	Il cluster 70	70	70	Movimentazione in 20000 contante del cluster sopra la soglia
Finanza islamica	Il cluster 90	90	90	Importo Bonifici 10000 Rischiosi
Finanza islamica	Il cluster 80	80	80	Numero Banconote 10 da 500 utilizzate
Finanza islamica	Il cluster 90	90	90	Importo Totale 0 Compravendite
Finanza islamica	Soggetti del 80	80	80	Numero soggetti nel 0 cluster in liste famiglie calabresi
Finanza islamica	Soggetti del 80	80	80	Numero soggetti nel 0 cluster in liste criminali
Finanza islamica	Soggetti del 80	80	80	Numero PEP o 0 criminali
Finanza islamica	Spese con 80	80	80	Importo spese e 1000 prelievi all'estero
Fondazioni Islamiche	Bonifici verso paesi a fiscalità agevolata	90	90	Importo Bonifici Totale 5000

Fondazioni Islamiche	Il cluster effettua Bonifici Esteri	80	80	80	Importo bonifici	totale 5000
Fondazioni Islamiche	Il cluster effettua bonifici per cassa	90	90	90	Importo Bonifici per cassa	10
Fondazioni Islamiche	Il cluster effettua movimenti contanti	90	90	90	Importo Movimenti	Totale 20000
Fondazioni Islamiche	Il cluster esegue Bon. rischiosi Importo > 10000	90	90	90	Importo Rischiosi	Bonifici 10000
Fondazioni Islamiche	Il cluster utilizza banconote da 500 eur	80	80	80	Numero Banconote da 500 utilizzate	10
Fondazioni Islamiche	Il cluster utilizza moneta virtuale	90	90	90	Importo Compravendite	Totale 0
Fondazioni Islamiche	Soggetti del cluster da famiglie calabresi	80	80	80	Numero soggetti nel cluster in liste famiglie calabresi	0
Fondazioni Islamiche	Soggetti del cluster in liste crimini	80	80	80	Numero soggetti nel cluster in liste criminali	0
Fondazioni Islamiche	Soggetti del cluster in liste pep	90	90	90	Numero di criminali o PEP nel cluster	0
Fondazioni Islamiche	Spese con carte all'estero	70	70	70	Importo totale spese e prelievi	1000
Hawalla	Bonifici verso paesi a fiscalità agevolata	90	90	90	Importo Bonifici	totale 5000
Hawalla	Il cluster effettua bonifici Esteri	80	80	80	Bonifici Esteri	10
Hawalla	Il cluster effettua bonifici per cassa	90	90	90	Importo Bonifici per cassa	10
Hawalla	Il cluster effettua	90	90	90	Importo Movimenti	Totale 20000

Hawalla	movimenti contanti	Il cluster 90	90	90	Importo Rischiosi	Bonifici 10000
Hawalla	esegue Bon. rischiosi Importo > 10000	Il cluster ha 80	80	80	Importo Movimenti	Totale 150000
Hawalla	SAE600 che muovono contanti	Il cluster 80	80	80	Numero Banconote 10 da 500 utilizzate	
Hawalla	utilizza banconote da 500 eur	Soggetti del 80	80	80	Numero soggetti nel 0 cluster in liste famiglie calabresi	
Hawalla	famiglie calabresi	Soggetti del 80	80	80	Numero soggetti nel 0 cluster in liste criminali	
Hawalla	cluster in liste criminali	Soggetti del 90	90	90	Numero di criminali o 0 PEP ammessi nel cluster	
Hawalla	pep Spese con 70	70	70	70	Totale spese e 1000 prelievi del cluster all'estero	
	carte all'estero					

9. WORKFLOW DI LAVORAZIONE

Il workflow di approvazione può essere personalizzato per ogni singolo cliente per adattarlo alle esigenze organizzative. Un esempio di wkfl base è il seguente:



10. GLOSSARIO DEI TERMINI UTILIZZATI

- **CURVE ROC (Receiver Operating Characteristic):** è lo schema grafico per un classificatore binario. Lungo i due assi si possono rappresentare la sensibilità e (1-specificità), come True Positive Rate (TPR, frazione di veri positivi) e False Positive Rate (FPR, frazione di falsi positivi).
- **ACCURATEZZA:** percentuale del numero di istanze classificate correttamente sul totale di istanze classificate.
- **ERRORE (1-accuratezza):** percentuale del numero di istanze classificate non corrette sul totale di istanze classificate.
- **Kappa di Kohen:** coefficiente statistico che rappresenta il grado di accuratezza e affidabilità in una classificazione statistica