

Manuale dell'utente di IBM SPSS
Modeler Entity Analytics 15



Nota: Prima di utilizzare queste informazioni e il relativo prodotto, leggere le informazioni generali disponibili in Note a pag. 65.

Questa versione si applica a e a tutte le successive versioni e modifiche fino a eventuali disposizioni contrarie indicate in nuove versioni.

Le schermate dei prodotti Adobe sono state ristampate su autorizzazione di Adobe Systems Incorporated.

Le schermate dei prodotti Microsoft sono state ristampate su autorizzazione di Microsoft Corporation.

Materiali concessi in licenza - Proprietà di IBM

Tutti i diritti riservati.

Prefazione

IBM® SPSS® Modeler è l'efficace workbench di data mining aziendale di IBM Corp.. SPSS Modeler consente alle organizzazioni di migliorare le relazioni con i clienti e con il pubblico grazie a un'analisi approfondita dei dati. Le organizzazioni potranno utilizzare le informazioni ottenute tramite SPSS Modeler per mantenere i clienti di valore, cogliere opportunità di vendite incrociate, attrarre nuovi clienti, individuare frodi, diminuire i rischi e migliorare l'offerta di servizi a livello statale.

L'interfaccia visiva di SPSS Modeler favorisce l'applicazione di una competenza aziendale specifica da parte degli utenti, grazie alla quale sarà possibile ottenere modelli di previsione più efficaci e una riduzione nei tempi di sviluppo delle soluzioni. SPSS Modeler offre una vasta gamma di tecniche di creazione di modelli, quali previsione, classificazione, segmentazione e algoritmi per l'individuazione delle associazioni. IBM® SPSS® Modeler Solution Publisher consente quindi di distribuire a livello aziendale i modelli creati in modo che vengano utilizzati dai responsabili dei processi decisionali oppure inseriti in un database.

Informazioni su IBM Business Analytics

Il software IBM Business Analytics fornisce informazioni complete, coerenti e accurate a cui i responsabili delle decisioni possono affidarsi per ottimizzare le prestazioni dell'azienda. Un ampio portafoglio di applicazioni di [business intelligence](#), [analisi predittiva](#), [gestione delle prestazioni e delle strategie finanziarie](#) e [analisi](#) offre una panoramica chiara, istantanea e interattiva delle prestazioni attuali e la possibilità di prevedere i risultati futuri. Utilizzato in combinazione con potenti soluzioni di settore, prassi consolidate e servizi professionali, questo software consente alle aziende di tutte le dimensioni di ottimizzare la produttività, automatizzare le decisioni senza problemi e fornire risultati migliori.

Come parte di questo portafoglio, il software IBM SPSS Predictive Analytics consente alle aziende di prevedere gli eventi futuri e di agire tempestivamente in modo da migliorare i risultati delle attività aziendali. Le aziende, gli enti governativi e le università di tutto il mondo si affidano alla tecnologia IBM SPSS perché rappresenta un vantaggio concorrenziale in termini di attrazione, retention e aumento dei clienti, riducendo al tempo stesso le frodi e limitando i rischi. Incorporando il software IBM SPSS nelle attività quotidiane, le aziende diventano imprese in grado di effettuare previsioni e di gestire e automatizzare le decisioni, per raggiungere gli obiettivi aziendali e vantaggi tangibili sulla concorrenza. Per ulteriori informazioni o per contattare un rappresentante, visitare il sito <http://www.ibm.com/spss>.

Supporto tecnico

Il supporto tecnico è a disposizione dei clienti che dispongono di un contratto di manutenzione. I clienti possono contattare il supporto tecnico per richiedere assistenza per l'utilizzo di IBM Corp. o per l'installazione di uno degli ambienti hardware supportati. Per contattare il supporto tecnico, visitare il sito Web IBM Corp. all'indirizzo <http://www.ibm.com/support>. Ricordare che durante la richiesta di assistenza sarà necessario fornire i dati di identificazione personali, i dati relativi alla propria società e il numero del contratto di manutenzione.

Contenuto

1	<i>Analisi delle entità</i>	1
	L'analisi delle entità	1
	Analisi delle entità e analisi predittiva	2
2	<i>Analisi delle entità con IBM SPSS Modeler</i>	4
	Utilizzo di Entity Analytics con IBM SPSS Modeler	4
	Fase 1: leggere i dati di origine in SPSS Modeler	5
	Fase 2: creare il repository	6
	Fase 3: connettere SPSS Modeler al repository	7
	Fase 4: mappare i campi di input alle funzioni del repository	7
	Fase 5: esportare i dati nel repository e risolvere i conflitti	8
	Fase 6: Analisi delle identità risolte	11
	Fase 7: risolvere i nuovi casi in base al repository	11
	Fase 8: generare avvisi	13
3	<i>Attività di analisi delle entità</i>	14
	Introduzione	14
	Impostazione di un repository di entità (nodo Esportazione EA)	14
	Il repository di entità	14
	Connessione a una sorgente dati	15
	Creazione del repository	16
	Mappatura dei campi di input alle funzioni (nodo Esportazione EA)	19
	Visualizzazione delle mappature dei campi (nodo Esportazione EA)	21
	Configurazione di un repository di entità	22
	Visualizzazione delle mappature delle sorgenti dati	23
	Gestione delle funzioni del repository	24
	Aggiunta o modifica di una funzione	26
	Gestione dei tipi di entità	28
	Impostazione della soglia per la corrispondenza delle entità	30
	Riutilizzo di una configurazione di repository	31
	Salvataggio delle modifiche alla configurazione	32
	Chiusura della finestra di configurazione	32
	Analisi delle identità risolte (nodo di input Entity Analytics(EA))	32
	Selezione di una sorgente dati	32
	Ridenominazione dei campi dati	33
	Impostazione delle informazioni sui tipi per i campi dati	34
	Aggiunta di nodi allo stream	34

Confronto di nuovi casi con il repository (nodo Streaming di EA)	34
Mappatura dei campi di input alle funzioni (nodo Streaming di EA)	36
Visualizzazione delle mappature dei campi (nodo Streaming di EA)	37
Visualizzazione delle sorgenti dati (nodo Streaming di EA)	38
Output del nodo Streaming di EA	39
Utilizzo di IBM SPSS Modeler Entity Analytics con altri prodotti IBM SPSS	39
Attività amministrative	40
Configurazione delle assegnazioni delle porte	40
Gestione delle credenziali dell'amministratore per il database del repository	41
Spostamento della directory di archiviazione del repository	42
Impostazione delle opzioni dello stream per i campi data/ora e timestamp	42
Esecuzione di IBM SPSS Modeler Entity Analytics con il client SPSS Modeler e SPSS Modeler Server sul medesimo sistema Windows	43
Svuotamento di un repository di entità	43
Eliminazione di un repository di entità	43
Eliminazione di un repository quando non è possibile connettersi a esso	44

4 Funzionamento di Entity Analytics 46

Informazioni su questo esempio	46
Il modello originale	46
Aggiunta dell'analisi delle entità	50
Inserimento dei dati di origine nel repository	51
Lettura delle identità risolte	52
Confronto dell'output dell'analisi delle entità con il modello originale	60
Riepilogo	64

Appendice

A Note 65

Indice 68

Analisi delle entità

L'analisi delle entità

IBM® SPSS® Modeler Entity Analytics aggiunge una dimensione completamente nuova alle analisi predittive di IBM® SPSS® Modeler. Se l'analisi predittiva tenta di prevedere il comportamento futuro sulla base di dati precedenti, l'analisi dell'entità si concentra sul miglioramento della coerenza dei dati correnti risolvendo i conflitti tra gli stessi record. Un'identità può essere di un individuo, un'organizzazione, un oggetto o qualsiasi altra entità per cui possa esistere ambiguità. La risoluzione dell'identità può essere essenziale in diversi campi, tra cui la gestione delle relazioni con i clienti, il rilevamento di frodi, il riciclaggio di denaro e la sicurezza nazionale e internazionale.

Si supponga di disporre dei seguenti record di clienti provenienti da due sorgenti diverse e di non sapere con esattezza se fanno riferimento alla stessa persona o a persone diverse.

Sorgente 1

Record n.: 70001
Nome: Jon Smith
Indirizzo: 123 Main Street
Riferimento imposta: 555-00-1111
Patente n.: 0001133107
Carta di credito n.: 10229127

Sorgente 2

Record n.: 9103
Nome: JOHNATHAN Smith
Data di nascita: 17/06/1934
Telefono: 555-1212
Carta di credito n.: 10229128
E-mail: jls@mail.com
Indirizzo IP: 9.50.18.77

Non esistono corrispondenze esatte dei dati tra i due record. Tuttavia, se si introduce una terza sorgente, si trovano alcuni attributi comuni.

Sorgente 3

Record n.: 6251
Nome: Jon Smith
Telefono: 555-1212
Patente n.: 0001133107
Carta di credito n.: 10229132

Il numero di patente collega i record della sorgente 1 e della sorgente 3, mentre il numero di telefono collega le sorgenti 2 e 3. Quindi si può essere ragionevolmente certi che tutte e tre le sorgenti si riferiscono alla stessa persona.

Ma cosa succede se la distinzione non è così semplice? I dati su cui basare la decisione potrebbero essere molto pochi. Consideriamo i seguenti due record.

Sorgente 4

Record n.: S45286
Nome: John T Smith Jr
Indirizzo: 456 Main Street
Telefono: 703-555-2000
Data di nascita: 12/03/1984

Record n.: S45287
Nome: John T Smith
Indirizzo: 456 Main Street
Telefono: 703-555-2000
Patente n.: 009900991

Evidentemente non si tratta dello stesso signor Smith dei record precedenti: le differenze sono sufficienti per stabilirlo. Tuttavia, rimane ancora un problema. Due record differenti, dalla stessa sorgente dati, sembrano essere correlati alla stessa persona. Sono record duplicati? Non è possibile saperlo con certezza se non si trova un altro record correlato che fornisca ulteriori informazioni, possibilmente da una sorgente diversa.

Sorgente 5

Record n.: 769582
Nome: Jon T Smith Sr
Indirizzo: 456 Main Street
Telefono: 703-555-2000
Patente n.: 009900991
Data di nascita: 25/06/1959

Questo risolve il problema. I due record nella sorgente 4 non sono duplicati, ma si riferiscono a padre e figlio con lo stesso nome, che abitano allo stesso indirizzo e utilizzano lo stesso numero di telefono. In un sistema manuale, possono essere necessarie settimane di ricerca per trovare l'unico record che risolve le identità. Con un sistema di analisi delle entità automatizzato, i tempi di risoluzione vengono notevolmente ridotti.

Analisi delle entità e analisi predittiva

Se tutti i dati provengono da un'unica sorgente di record completi e privi di ambiguità, è relativamente semplice per IBM® SPSS® Modeler risolvere eventuali conflitti di identità. Utilizzando solo l'analisi predittiva, è possibile leggere i dati in SPSS Modeler, effettuare l'elaborazione e ottenere risultati affidabili.

Nella realtà, tuttavia, la situazione di solito è molto diversa. Normalmente i dati sono tutt'altro che completi, spesso sono ambigui e distribuiti in svariate sorgenti dati e vengono registrati molti attributi diversi con pochi campi che si sovrappongono. Parte del valore dell'analisi delle entità risiede nella raccolta dei dati da tutte le sorgenti dati in un'unica area di archiviazione centrale, nota come **repository**. Il sistema di analisi delle entità esamina quindi i dati in dettaglio per risolvere i conflitti, aggiungendo un identificatore univoco ai record che hanno origine dalla stessa persona o organizzazione.

Nella tabella che segue sono illustrate le differenze tra i due tipi di analisi.

Tabella 1-1

Differenze tra l'analisi predittiva e l'analisi delle entità

Caratteristica	Analisi predittiva	Analisi delle entità
Addestramento	Basato su insiemi e intervalli numerici relativamente piccoli	Basato su insiemi di grandi dimensioni, ad esempio nomi, indirizzi
Dati utilizzati per l'addestramento	Sottoinsieme (partizione addestramento)	Tutti i dati utilizzati
Generalizzazione	L'algoritmo generalizza i dati dell'addestramento per creare un modello conciso	I dati rimangono in strutture adatte per la corrispondenza delle entità e il rilevamento delle relazioni
Rilevamento di frodi	I record vengono contrassegnati come potenzialmente fraudolenti se presentano le caratteristiche tipiche dell'applicazione fraudolenta	I record vengono contrassegnati come potenzialmente fraudolenti se correlati a record sicuramente fraudolenti o se hanno origine dagli stessi individui con identità diverse

Analisi delle entità con IBM SPSS Modeler

Utilizzo di Entity Analytics con IBM SPSS Modeler

Si ritiene di poter avere problemi di identità con i dati in uso. In che modo IBM® SPSS® Modeler Entity Analytics può aiutare a risolverli? Di seguito viene fornito un suggerimento di procedura, che potrà essere modificata in modo da adattarsi ai requisiti particolari.

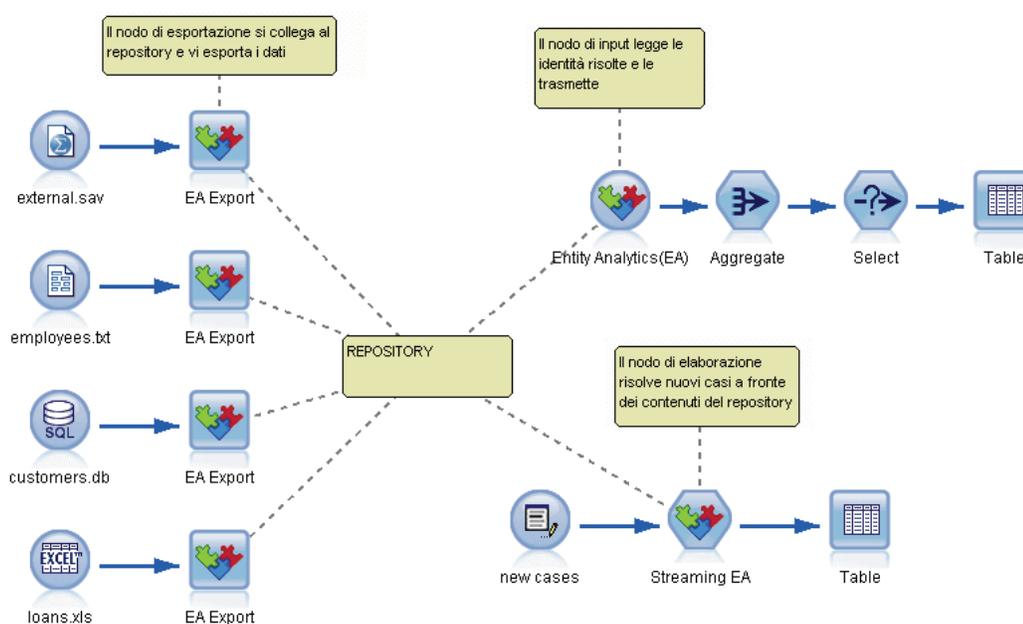
- leggere i dati di origine in IBM® SPSS® Modeler
- Creare un repository pronto per l'archiviazione dei dati
- connettere SPSS Modeler al repository
- Mappare i campi dati alle funzioni del repository
- Esportare i dati nel repository e risolvere le identità
- Analizzare le identità risolte
- risolvere i nuovi casi in base al repository
- Generare eventuali avvisi necessari (batch o tempo reale)

A questo punto, sono necessarie alcune informazioni sul funzionamento di SPSS Modeler. SPSS Modeler è uno strumento di facile uso, basato sulla rappresentazione grafica di uno stream di dati che passa attraverso un numero di nodi. Ogni nodo rappresenta una fase particolare del flusso di lavoro.

SPSS Modeler offre un'ampia gamma di nodi, che coprono tutte le funzioni standard di data mining. IBM SPSS Modeler Entity Analytics aggiunge nodi da utilizzare specificamente nell'analisi delle entità. Si tratta del nodo Esportazione EA, del nodo di input Entity Analytics(EA) e del nodo di elaborazione Streaming di EA.

Il processo è illustrato nella figura seguente.

Figura 2-1
Il processo di analisi delle entità



Fase 1: leggere i dati di origine in SPSS Modeler

La prima operazione è leggere i dati in SPSS Modeler per mezzo di un nodo di input, indicato in SPSS Modeler con un'icona rotonda.

Figura 2-2
nodo Input IBM SPSS Modeler



I dati possono essere in qualsiasi formato supportato da SPSS Modeler, ad esempio file di testo, tabelle di database, fogli di calcolo, file XML e così via, ma ogni diverso formato richiede un nodo di input di SPSS Modeler corrispondente. Nell'illustrazione è un nodo di input Database.

Ogni file della sorgente dati deve avere un campo che identifichi in modo univoco ogni record. Se una sorgente dati non contiene tale campo, è possibile aggiungerne facilmente uno in SPSS Modeler. Per ulteriori informazioni, vedere l'argomento [Aggiunta di un identificatore univoco del record](#) in il capitolo 3 a pag. 15.

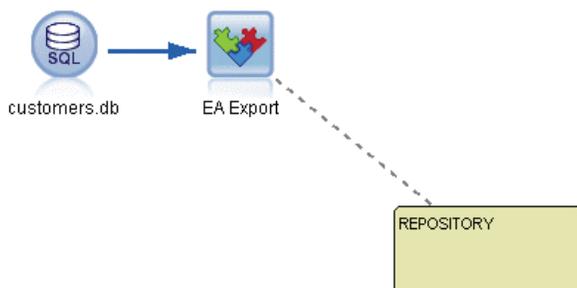
Per ulteriori informazioni, vedere l'argomento [Connessione a una sorgente dati](#) in il capitolo 3 a pag. 15.

Fase 2: creare il repository

Il punto focale di tutte le attività di analisi delle entità è il repository, l'area di archiviazione centrale in cui vengono raccolti tutti i record di dati.

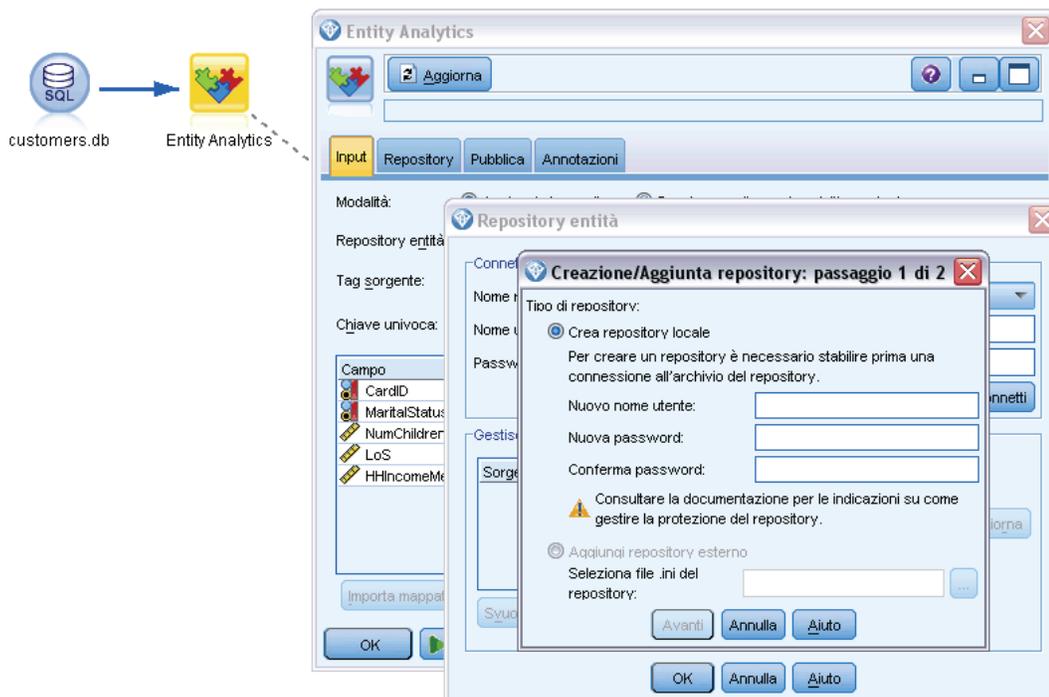
Per creare un repository, per prima cosa si connette la sorgente dati a un nodo Esportazione EA, rappresentato con un'icona quadrata.

Figura 2-3
Connessione a un repository



Dal nodo di esportazione è possibile creare un nuovo repository (o selezionarne uno esistente), pronto per ricevere i dati esportati.

Figura 2-4
Creazione di un repository



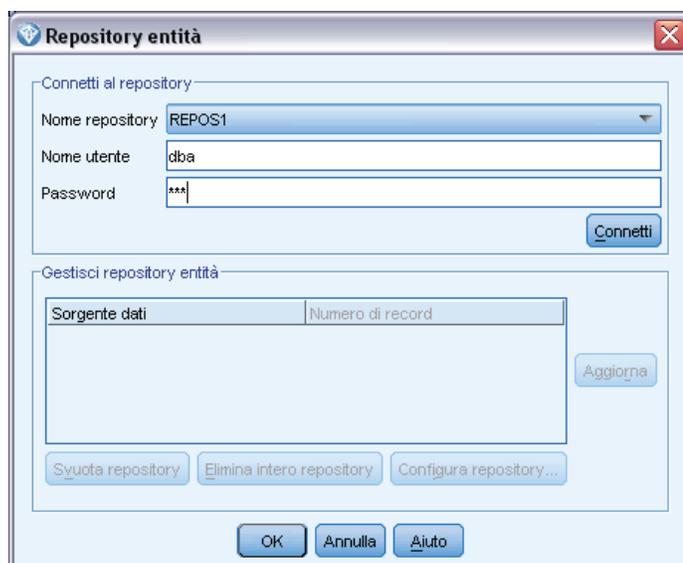
Il processo di creazione di un repository è descritto in dettaglio più avanti. Per ulteriori informazioni, vedere l'argomento [Impostazione di un repository di entità \(nodo Esportazione EA\)](#) in il capitolo 3 a pag. 14.

Una volta impostato un repository, è possibile gestirne il contenuto in vari modi. Per ulteriori informazioni, vedere l'argomento [Configurazione di un repository di entità](#) in il capitolo 3 a pag. 22.

Fase 3: connettere SPSS Modeler al repository

Dopo avere creato il repository, lo si connette allo stream di SPSS Modeler.

Figura 2-5
Connessione al repository



Per ulteriori informazioni, vedere l'argomento [Opzioni del repository di entità](#) in il capitolo 3 a pag. 18.

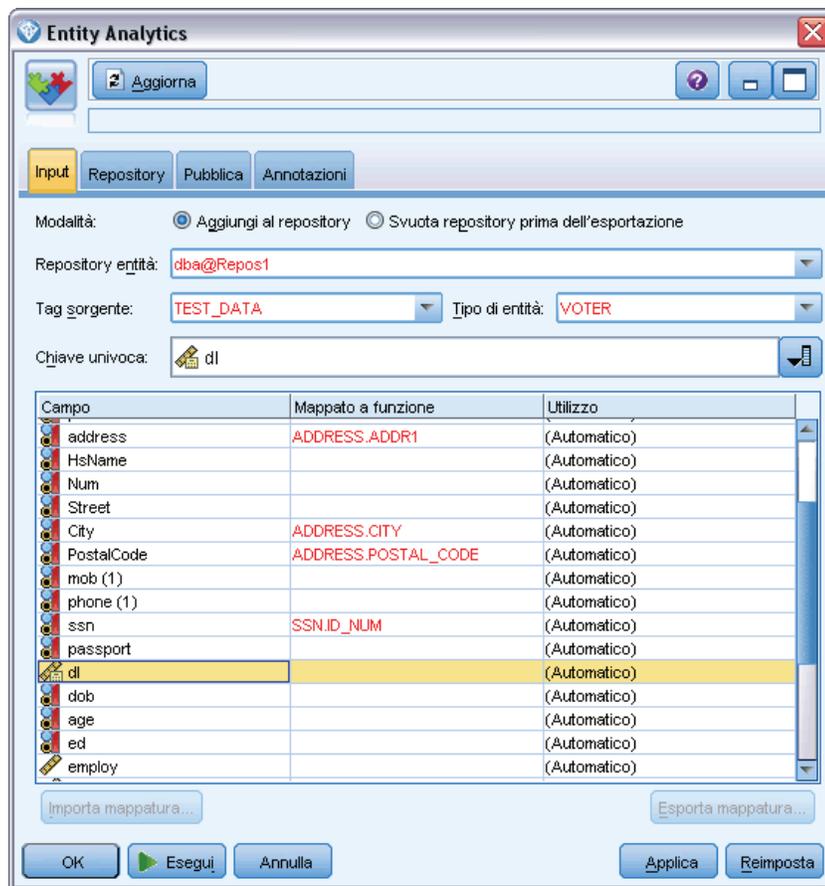
Fase 4: mappare i campi di input alle funzioni del repository

Le sorgenti dati possono contenere molti tipi diversi di informazioni sulle entità. Alcuni tipi di informazioni sono comuni alla maggior parte delle sorgenti dati delle entità, mentre altri possono essere specifici di una particolare sorgente dati. In un repository di entità i diversi tipi di informazioni sono detti **funzioni**. Il repository fornisce diverse funzioni come standard ed è anche possibile creare le proprie funzioni.

Una funzione di repository è un singolo tipo di informazione che può essere utilizzato con una sorgente dati di entità. Alcune funzioni (per esempio, Nome, Cognome, Data di nascita e così via) possono essere utilizzate da molte sorgenti dati diverse, mentre altre funzioni sono specifiche di una sorgente dati particolare. Una funzione normalmente equivale a un campo di un record di dati o a una colonna in una tabella di database.

Una volta creato un repository ed eseguita la connessione, si designa un campo dei dati di input come **campo chiave univoco**, che verrà utilizzato nell'analisi successiva. Inoltre, i campi dei dati di input devono essere mappati alle funzioni corrispondenti nel repository, per evitare duplicati nel repository nel caso in cui sorgenti dati diverse utilizzino nomi diversi per i campi contenenti lo stesso tipo di informazioni. Il nodo Esportazione EA fornisce una tabella di mappatura in cui è possibile creare le mappature.

Figura 2-6
Mappatura dei campi di input alle funzioni

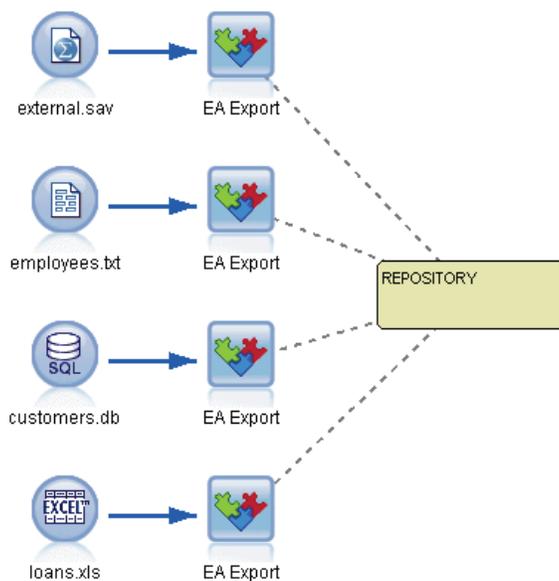


Per ulteriori informazioni, vedere l'argomento [Mappatura dei campi di input alle funzioni \(nodo Esportazione EA\)](#) in il capitolo 3 a pag. 19.

Fase 5: esportare i dati nel repository e risolvere i conflitti

Ogni nodo di sorgente dati richiede un proprio nodo Esportazione EA, quindi se i dati vengono distribuiti in diverse sorgenti, è possibile che risulti uno stream come questo.

Figura 2-7
Esportazione dei dati nel repository da più sorgenti dati



È possibile scegliere di leggere i record da una, alcune o tutte le sorgenti dati se si dispone di più sorgenti. Il sistema Entity Analytics analizza i record selezionati e aggiunge il campo identificatore \$EA_ID a ognuno di essi. Se due o più record correlati a identità in precedenza ambigue ora possono essere risolti, gli identificatori aggiunti a tali record sono univoci in tutto il repository. Il sistema aggiunge anche un campo per mostrare la sorgente dati da cui hanno origine i record.

È necessario connettere ciascun nodo di sorgente dati al proprio nodo Esportazione EA, mappare i campi di input alle funzioni del repository e quindi eseguire lo stream per esportare i dati da SPSS Modeler al repository e risolvere eventuali conflitti di identità, in un'unica operazione. Per vedere come funziona, si supponga di disporre dei seguenti record in quattro sorgenti dati diverse.

Dati esterni

Nome	Telefono	Rischio credito
Mike	555-1234	560
Joe	555-4567	780

Dipendenti

Nome	Indirizzo	Telefono
Michael	1234 5th Street	555-1234
Fred	543 1st Avenue	555-9876

Clienti

Nome	Indirizzo	Risparmio
Susan	1234 5th Street	\$1234
Joe	777 Oak Street	\$5

Prestiti

Nome	Indirizzo	Telefono	Prestito
Sue	1234 5th Street	555-1234	\$10,000
Joseph	777 Oak Street	555-4567	\$50,000

Come si può vedere, si esportano ad una ad una tutte le sorgenti dati nel repository. In questo modo, il repository aggiorna la risoluzione di ogni record. Nel repository, ciascun record è preceduto da un campo identificatore (denominato *SEA-ID*) e un campo indicatore della sorgente (denominato *SEA-SRC*), che indica la sorgente dati da cui ha origine il record. Quindi, in questo esempio, una volta esportate tutte e quattro le sorgenti dati, il contenuto del repository è simile a questo.

Tabella 2-1

Esempio di contenuto del repository dopo la fase di esportazione

<i>SEA-ID</i>	<i>SEA-SRC</i>	Nome	Telefono	Indirizzo	Rischio credito	Risparmio	Prestito
1	Dipendenti	Michael	555-1234	1234 5th St			
1	External	Mike	555-1234		560		
2	Clienti	Joe		777 Oak St		\$5	
2	External	Joe	555-4567		780		
2	Prestiti	Joseph	555-4567	777 Oak St			\$50,000
3	Dipendenti	Fred	555-9876	543 1st Ave			
4	Clienti	Susan		1234 5th St		\$1234	
4	Prestiti	Sue	555-1234	1234 5th St			\$10,000

Il sistema Entity Analytics ha determinato che *Mike* nell'insieme di dati *Esterni* è la stessa persona di nome *Michael* nell'insieme di dati *Dipendenti* in base a un numero di telefono in comune, e ha assegnato l'ID 1.

Il caso di *Joe* nell'insieme di dati *Esterni* è un po' più difficile. È la stessa persona di nome *Joe* in *Clienti*? È impossibile stabilirlo utilizzando solo due sorgenti dati, ma si dispone di una terza sorgente, *Prestiti*, che contiene un *Joseph*. Ora esiste una corrispondenza: il numero di telefono di *Joseph* è lo stesso di *Joe* nell'insieme di dati *Esterni*. Su questa base, il sistema determina che si tratta della stessa persona e assegna l'identificatore 2.

Non esistono record multipli per *Fred*, quindi viene assegnato l'ID 3. *Susan* di *Clienti* viene identificata come la *Sue* di *Prestito*, poiché hanno lo stesso indirizzo, quindi le viene assegnato l'ID 4.

Nota: questo esempio di corrispondenza ottimista è puramente a scopo illustrativo. Si può scegliere un insieme di regole più pessimista in modo che un semplice nome e numero di telefono o un indirizzo da solo non venga considerato una corrispondenza esatta che assegnerebbe ad entrambi i record il medesimo identificatore.

Fase 6: Analisi delle identità risolte

Dopo avere risolto i conflitti di identità nel repository, è possibile eseguire ulteriori analisi ed elaborazioni sui risultati. Ad esempio, se si sospetta una possibile attività fraudolenta correlata all'esistenza di record duplicati per la stessa identità, è possibile generare un report con l'elenco dei duplicati.

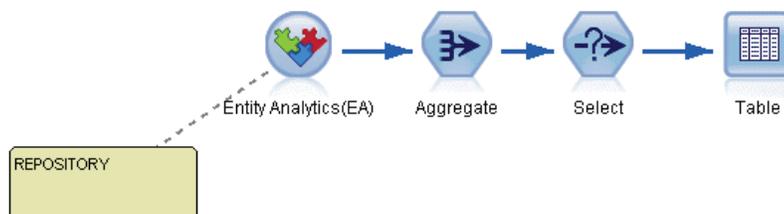
Si inizia creando un nodo di input Entity Analytics(EA) da collegare al repository.

L'output del nodo è costituito dai seguenti campi.

- Il campo identificatore aggiunto dal sistema (*\$EA-ID* nell'esempio di fase 5)
- Il campo indicatore della sorgente aggiunto dal sistema (*\$EA-SRC* nell'esempio di fase 5)
- Il campo chiave univoco definito nella fase 4.

Per visualizzare l'output in SPSS Modeler, è possibile collegare un nodo di output di SPSS Modeler, ad esempio un nodo Tabella o un nodo Report, ed eseguire questa parte dello stream. Se è necessario riassumere l'output, che potrebbe essere molto lungo, è possibile includere nodi di operazioni sui record come Aggregazione o Seleziona, come nell'illustrazione che segue.

Figura 2-8
Visualizzazione dell'output



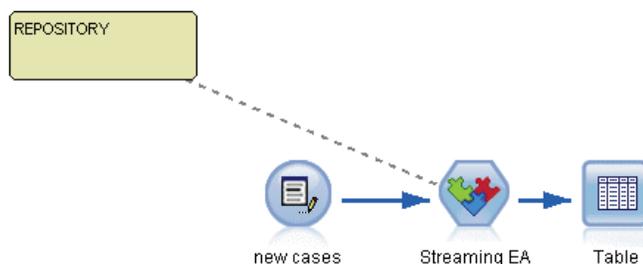
Il nodo di input Entity Analytics(EA) è descritto in dettaglio più avanti. Per ulteriori informazioni, vedere l'argomento [Analisi delle identità risolte \(nodo di input Entity Analytics\(EA\)\)](#) in il capitolo 3 a pag. 32.

Fase 7: risolvere i nuovi casi in base al repository

Le identità di tutti i record in tutte le sorgenti dati sono state risolte. Cosa accade quando si vuole confrontare un insieme di nuovi record per vedere come sono correlati a qualcosa che già conosciamo, per ottenere un punteggio migliore? Qui interviene il nodo Streaming di EA.

Per prima cosa, si aggiunge un nuovo nodo di input dei dati di SPSS Modeler per leggere i nuovi dati nello stream. Quindi, si connette il nodo di input a un nodo Streaming di EA. Per visualizzare l'output, si aggiunge un nodo Tabella come in precedenza, in modo tale che questa parte dello stream abbia il seguente aspetto.

Figura 2-9
Risoluzione dei nuovi casi



Quando si esegue questa parte dello stream, il nodo Streaming di EA legge ogni nuovo record e lo confronta con il contenuto del repository. Se rileva dei record corrispondenti nel repository, il nodo Streaming di EA li indica nell'output insieme al nuovo record, al quale vengono aggiunti i campi ID e indicatore di sorgente. Se non vengono rilevate corrispondenze, l'output del nodo di elaborazione contiene solo il nuovo record con i campi ID e indicatore di sorgente.

Per esempio, si supponga che il repository attualmente sia costituito dai contenuti generati dal nodo di input Entity Analytics(EA). Vedere [Tabella 2-1](#) a pag. 10.

Si ricevono i seguenti nuovi record. Sono correlati a qualcuno che già conosciamo?

Tabella 2-2
Nuovi record da calcolare

Nome	Indirizzo	Telefono	Prestito
Suzan	1234 5th Street	555-1234	\$100,000
Mark	888 9th Ave	555-9999	\$60,000

Confrontando i nuovi dati con il contenuto del repository già esistente, il nodo Streaming di EA mette in relazione il primo nuovo record con la persona che presenta l'identificatore 4 nei record esistenti. Tuttavia, non vengono trovate corrispondenze per il secondo nuovo record, a cui viene assegnato un nuovo identificatore univoco, 5.

Il nodo Streaming di EA aggiunge i campi identificatore e indicatore della sorgente e genera un output con i nuovi record e tutti i relativi record corrispondenti. Quindi l'output sarà come segue.

Tabella 2-3
Output del nodo Streaming di EA

SEA-ID	SEA-SRC	Nome	Telefono	Indirizzo	Rischio credito	Risparmio	Prestito
4	Cliente	Susan		1234 5th St		\$1234	
4	Prestito	Sue	555-1234	1234 5th St			\$10,000

SEA-ID	SEA-SRC	Nome	Telefono	Indirizzo	Rischio credito	Risparmio	Prestito
4	Nuovo prestito	Suzan	555-1234	1234 5th Street			\$100,000
5	Nuovo prestito	Mark	555-9999	888 9th Ave			\$60,000

Questo output può quindi essere aggregato in base agli identificatori dell'analisi delle entità e passato ai nodi a valle per un'ulteriore elaborazione.

Il nodo Streaming di EA è descritto in dettaglio più avanti.

Fase 8: generare avvisi

Ancora una volta, potrebbero essere rilevate attività potenzialmente sospette. In questo caso, la persona con identificatore 4 ha già un prestito di \$10.000 e sta richiedendo, utilizzando un nome leggermente diverso, un altro prestito per un importo dieci volte superiore. Questo naturalmente può essere perfettamente accettabile e non nascondere alcuna intenzione fraudolenta. Tuttavia, se tale attività viene considerata sospetta dalle proprie regole aziendali, vale la pena approfondire la questione.

È possibile, ad esempio, collegare ed eseguire un nodo Tabella o un nodo Report di SPSS Modeler, stampare il contenuto della finestra di output, leggerlo e generare manualmente gli avvisi. In alternativa, è possibile passare l'output al nodo Streaming di EA per un modello di valutazione del rischio creato in precedenza in IBM® SPSS® Modeler, per generare una serie di punteggi che riflettano meglio le regole aziendali. Un'altra possibilità è esportare l'output in un database o un altro tipo di supporto per un'ulteriore elaborazione. SPSS Modeler mette a disposizione un'ampia gamma di possibilità per soddisfare esigenze particolari.

Attività di analisi delle entità

Introduzione

Questa sezione descrive le seguenti attività di analisi delle entità.

- Impostazione di un repository di entità
- Configurazione di un repository di entità
- Analisi delle identità risolte
- Risoluzione di nuovi casi in base al repository di entità
- Svuotamento di un repository di entità
- Eliminazione di un repository di entità
- Amministrazione dell'analisi delle entità

Impostazione di un repository di entità (nodo Esportazione EA)

Il processo di impostazione di un repository di entità è costituito dalle seguenti operazioni.

1. Connessione a una sorgente dati Per ulteriori informazioni, vedere l'argomento [Connessione a una sorgente dati](#) a pag. 15.
2. Creazione del repository. Per ulteriori informazioni, vedere l'argomento [Creazione del repository](#) a pag. 16.
3. Mappatura dei campi di input nella sorgente dati alle funzioni del repository. Per ulteriori informazioni, vedere l'argomento [Mappatura dei campi di input alle funzioni \(nodo Esportazione EA\)](#) a pag. 19.

Una volta impostate le mappature, è possibile visualizzarle per la sorgente dati corrente o per tutte le sorgenti dati conosciute nel repository. Per ulteriori informazioni, vedere l'argomento [Visualizzazione delle mappature dei campi \(nodo Esportazione EA\)](#) a pag. 21.

Il repository di entità

Il repository fornisce un'area di archiviazione centrale, che funziona come cache di dati per tutte le informazioni relative alle entità. Poiché le attività del repository sono in tempo reale, il repository di entità ha un solo stato e non esiste alcun controllo delle versioni. Il repository mantiene lo stato corrente di tutti i dati di input, quindi può diventare molto grande.

È possibile gestire il contenuto del repository mediante un'interfaccia grafica di facile uso. Per ulteriori informazioni, vedere l'argomento [Configurazione di un repository di entità](#) a pag. 22.

Nota: La versione di IBM® SPSS® Modeler Entity Analytics fornita con IBM® SPSS® Modeler Premium supporta un unico repository ospitato sul prodotto IBM solidDB incluso con IBM SPSS Modeler Entity Analytics. Con questa versione, è necessario eliminare un repository esistente prima di crearne uno nuovo. È disponibile un aggiornamento con licenza separata di IBM SPSS Modeler Entity Analytics (noto come IBM SPSS Modeler Entity Analytics Unleashed) che ammette la coesistenza di più di un repository sul medesimo sistema; ogni repository può contenere oltre 10 milioni di righe e utilizzare più di quattro core del processore. Per maggiori dettagli, rivolgersi al proprio addetto del supporto IBM.

Connessione a una sorgente dati

Si inizia leggendo i dati di origine in SPSS Modeler utilizzando un nodo di input.

Per eseguire la connessione a una sorgente dati

1. Dalla scheda Input della palette dei nodi nella parte inferiore della finestra principale di SPSS Modeler, fare doppio clic su un'icona corrispondente al tipo di dati di origine. In questo modo si aggiunge un nodo Input all'area di disegno visualizzata.
2. Nell'area di disegno visualizzata, fare doppio clic sull'icona per aprire la relativa finestra di dialogo.
3. Inserire nel campo File il percorso e il nome del file di dati di origine.
4. Completare il resto della finestra di dialogo, se necessario (fare clic sul pulsante Aiuto per ulteriori informazioni), quindi fare clic su OK.
5. Se il file di dati di origine non contiene un campo che identifica in modo univoco ogni record, aggiungerne uno mediante un nodo Nuovo campo. Per ulteriori informazioni, vedere l'argomento [Aggiunta di un identificatore univoco del record](#) a pag. 15.

Aggiunta di un identificatore univoco del record

Ogni file della sorgente dati che viene inserito nel repository di entità deve avere un campo che identifichi in modo univoco ogni record. Se il file della sorgente dati non contiene tale campo, è possibile aggiungerne uno mediante un nodo Nuovo campo di SPSS Modeler.

Per aggiungere un identificatore univoco del record a un file della sorgente dati

1. Nell'area di disegno visualizzata, fare clic sul nodo Input aggiunto nell'operazione precedente.
2. Dalla scheda Oper su campi della palette dei nodi, fare doppio clic sull'icona Nuovo campo per collegare un nodo Nuovo campo al nodo Input.
3. Nell'area di disegno visualizzata, fare doppio clic sul nodo Nuovo campo per aprire la relativa finestra di dialogo.
4. Nel campo Nuovo campo, sostituire il nome di default con un nome significativo (ad esempio, ID) per il campo identificatore che si sta aggiungendo.

5. Assicurarsi che il campo Deriva come sia impostato su Formula
6. Impostare Tipo campo su Continuo.
7. Nella casella di testo Formula digitare @INDEX e fare clic su OK.

Creazione del repository

È necessario creare un repository per archiviare tutti i dati di input.

Per creare un repository

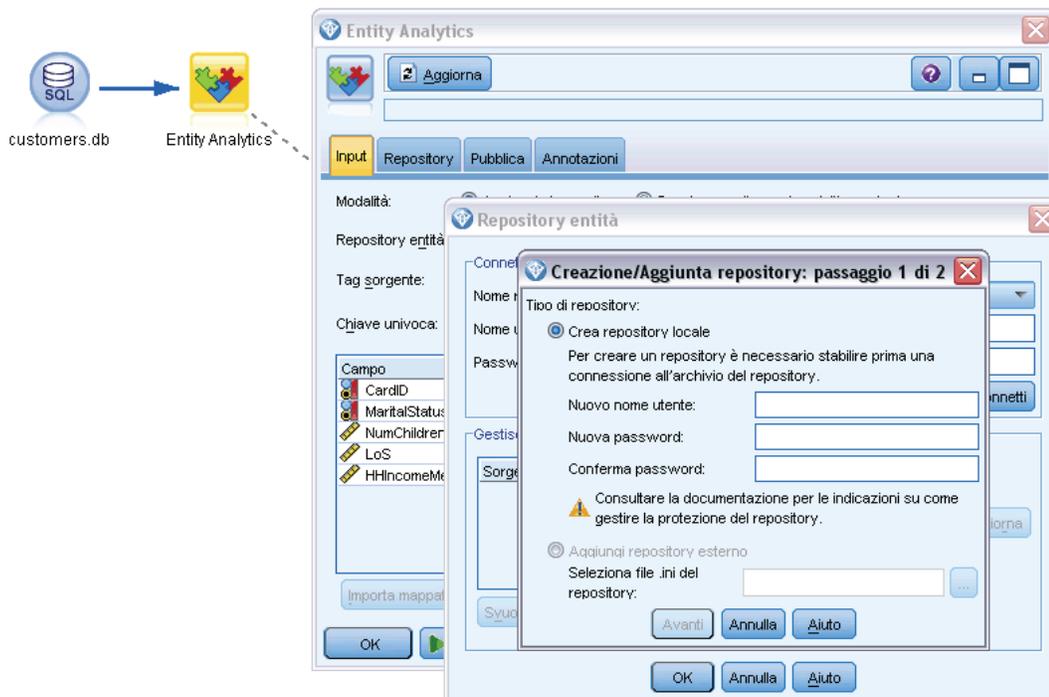
1. Dalla scheda Esporta della palette dei nodi di SPSS Modeler, posizionare un nodo Esportazione EA nell'area di disegno dello stream.

Nota: se si sta creando un repository per la prima volta, utilizzare un nodo Esportazione EA e connetterlo al nodo di input di SPSS Modeler che contiene i dati da inserire nel repository (oppure al nodo Nuovo campo, se ne è stato aggiunto uno per ottenere un campo identificatore univoco). Per connettere i nodi, effettuare le seguenti operazioni.

- Fare clic con il pulsante destro del mouse sul nodo di input di SPSS Modeler.
 - Scegliere Connetti.
 - Fare clic sul nodo Esportazione EA.
2. Fare doppio clic sul nodo Esportazione EA per aprire la relativa finestra di dialogo.
 3. Fare clic sull'elenco Repository entità.
 4. Fare clic su <Cerca...> per visualizzare la finestra di dialogo Repository entità.
 5. Nella finestra di dialogo Repository entità, fare clic sul campo Nome repository.
 6. Scegliere <Crea nuovo repository...> per visualizzare la procedura di creazione guidata del repository.

Creazione guidata del repository

Figura 3-1
Creazione guidata del repository



Fase 1

Qui si sceglie se creare un repository locale, mediante il prodotto IBM solidDB incluso con IBM SPSS Modeler Entity Analytics, oppure se utilizzare un database esterno per il repository.

Creazione repository locale. Specificare il nome utente e la password dell'amministratore per il database IBM solidDB che ospiterà il repository da creare. Confermare la password e fare clic su Avanti.

Nota: se successivamente è necessario modificare le credenziali dell'amministratore, è possibile farlo utilizzando l'editor della riga di comando per il database. Per ulteriori informazioni, vedere l'argomento [Gestione delle credenziali dell'amministratore per il database del repository](#) a pag. 41.

Aggiungi repository esterno. Utilizzare questa opzione per ospitare il repository su un database esterno. Digitare il percorso del file .ini del database nel campo Seleziona file .ini del repository e fare clic su Avanti.

Fase 2

Nuovo nome repository. Digitare un nome univoco per il nuovo repository.

Importa configurazione da. (Solo per repository locale) Se si desidera basare la configurazione su quella di un repository già esistente, scegliere qui il repository oppure scegliere Default. Per ulteriori informazioni, vedere l'argomento [Configurazione di un repository di entità](#) a pag. 22.

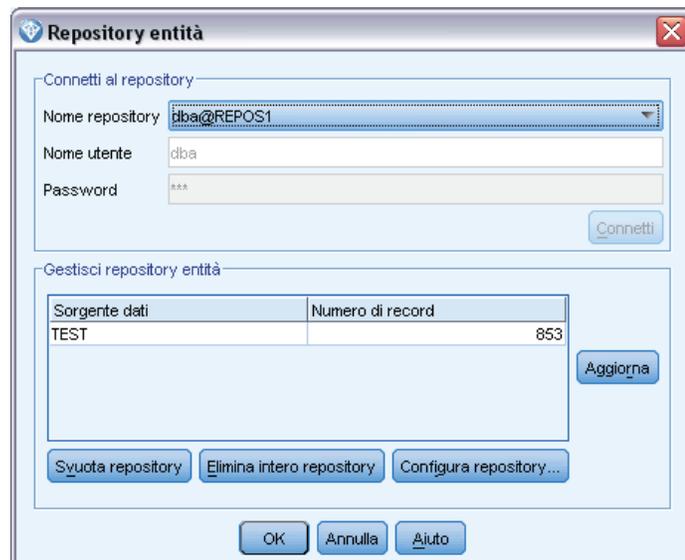
Se si sceglie un repository già esistente, inserire i dettagli della connessione se sono diversi da quelli immessi nella schermata precedente.

Fare clic su OK per creare il nuovo repository e visualizzare la finestra di dialogo Istanze di risoluzione entità, in cui è possibile eseguire la connessione al repository.

Opzioni del repository di entità

La finestra di dialogo Repository entità contiene alcune opzioni per la creazione, la connessione, la configurazione e la gestione del repository di entità.

Figura 3-2
Opzioni del repository di entità



Connetti al repository. Utilizzare queste opzioni per creare un nuovo repository di entità o per connettersi a un repository esistente.

- **Nome repository.** Visualizza il repository di entità corrente, se esiste. Per scegliere un repository diverso nel caso in cui ne esista più di uno, selezionarlo dall'elenco.
Per creare un nuovo repository, selezionare <Crea nuovo repository...>. In questo modo viene avviata una procedura di creazione guidata.
- **Nome utente.** Immettere il proprio nome utente per il repository selezionato.
- **Password.** Password del nome utente.
- **Connetti.** Selezionare per connettersi al repository corrente.

Gestisci repository entità. La tabella contiene le sorgenti dati che sono state caricate nel repository corrente (quello a cui si è connessi) e visualizza il numero di record in ogni sorgente dati.

- **Aggiorna.** Consente di aggiornare le informazioni sulla sorgente dati e sulle dimensioni nella tabella, ad esempio quando si aggiunge una nuova sorgente dati o si modificano le dimensioni di una sorgente già esistente.
- **Svuota repository.** Rimuove tutti i dati di origine dal repository mantenendo tutti i dettagli della configurazione. Questa opzione può essere utilizzata quando i dati di configurazione sono ancora utili ma si desidera rimuovere tutti i record di dati dal repository. Per ulteriori informazioni, vedere l'argomento [Svuotamento di un repository di entità](#) a pag. 43.

Elimina intero repository. Cancella completamente i contenuti del repository corrente e i dettagli della configurazione. Per ulteriori informazioni, vedere l'argomento [Eliminazione di un repository di entità](#) a pag. 43.

Configura repository. Visualizza una finestra in cui è possibile configurare il repository corrente. Per ulteriori informazioni, vedere l'argomento [Configurazione di un repository di entità](#) a pag. 22.

Mappatura dei campi di input alle funzioni (nodo Esportazione EA)

Il repository fornisce alcune funzioni predefinite come standard. Sorgenti dati diverse possono utilizzare nomi di campo diversi, ad esempio Indirizzo1 o Riga indirizzo 1, per i tipi di informazioni che corrispondono alla stessa funzione. Per evitare duplicati, è necessario mappare i campi di input della sorgente dati a funzioni specifiche del repository. Non è necessario mappare ogni campo nell'insieme di dati, sono sufficienti quelli che verosimilmente corrisponderanno alla stessa funzione in altri insiemi di dati.

Se una sorgente dati utilizza campi corrispondenti ad altri tipi di informazioni che non sono predefiniti nel repository, è possibile creare nuove funzioni dalla finestra Configurazione repository. Per ulteriori informazioni, vedere l'argomento [Configurazione di un repository di entità](#) a pag. 22.

Per mappare i campi di input alle funzioni

1. Collegare un nodo Esportazione EA a un nodo di input dei dati nell'area di disegno dello stream. Ogni nodo di input dei dati utilizzato deve essere collegato al proprio nodo Esportazione EA.
2. Aprire il nodo Esportazione EA per visualizzare la scheda Input, che contiene le opzioni di mappatura dei campi di input. Per ulteriori informazioni, vedere l'argomento [Opzioni di input del repository per la mappatura](#) a pag. 20.
3. Nel nodo Esportazione EA, selezionare la scheda Repository per visualizzare le assegnazioni della mappatura per la sorgente dati corrente o per tutte le sorgenti dati nel caso in cui se ne utilizzi più di una.
4. Per salvare un insieme di assegnazioni della mappatura (ad esempio, per utilizzarle con un altro nodo di input dei dati), fare clic su Esporta mappatura.

Dopo avere mappato il primo nodo di input dei dati, ripetere la procedura per tutti gli altri nodi di input dei dati che si desidera utilizzare.

Opzioni di input del repository per la mappatura

La scheda Input contiene le opzioni per la mappatura dei campi di input dei dati alle funzioni del repository pronte per l'esportazione nel repository. Impostare le assegnazioni della mappatura in questa scheda; se lo si desidera, fare clic sulla scheda Repository per visualizzare la mappatura delle altre sorgenti dati, quindi fare clic su Esegui per esportare i dati nel repository.

Se si è già archiviato un insieme di mappature in un file XML, è possibile utilizzarle facendo clic su Importa mappatura.

Figura 3-3
Mappatura dei campi di input alle funzioni

The screenshot shows the 'Entity Analytics' application window with the 'Input' tab selected. The interface includes a toolbar with an 'Aggiorna' button, a search bar, and navigation tabs for 'Input', 'Repository', 'Pubblica', and 'Annotazioni'. Below the tabs, there are configuration options for 'Modalità' (radio buttons for 'Aggiungi al repository' and 'Svuota repository prima dell'esportazione'), 'Repository entità' (dropdown menu showing 'dba@Repos1'), 'Tag sorgente' (dropdown menu showing 'TEST_DATA'), 'Tipo di entità' (dropdown menu showing 'VOTER'), and 'Chiave univoca' (text field showing 'di').

Campo	Mappato a funzione	Utilizzo
address	ADDRESS.ADDR1	(Automatico)
HsName		(Automatico)
Num		(Automatico)
Street		(Automatico)
City	ADDRESS.CITY	(Automatico)
PostalCode	ADDRESS.POSTAL_CODE	(Automatico)
mob (1)		(Automatico)
phone (1)		(Automatico)
ssn	SSN.ID_NUM	(Automatico)
passport		(Automatico)
di		(Automatico)
dob		(Automatico)
age		(Automatico)
ed		(Automatico)
employ		(Automatico)

At the bottom of the window, there are buttons for 'Importa mappatura...', 'Esporta mappatura...', 'OK', 'Esegui', 'Annulla', 'Applica', and 'Reimposta'.

Modalità. Mantenere la selezione di default, Aggiungi al repository, per aggiungere i record del file sorgente al contenuto del repository esistente. Per cancellare il contenuto del repository pur mantenendo le informazioni di configurazione prima di aggiungere i record sorgente, scegliere Svuota repository prima dell'esportazione.

Repository entità. Visualizza il repository di entità corrente, se esiste. Per scegliere un repository diverso nel caso in cui ne esista più di uno, selezionarlo dall'elenco. Per creare un nuovo repository, scegliere <Cerca...> per visualizzare una finestra di dialogo da cui è possibile creare

il repository. Per ulteriori informazioni, vedere l'argomento [Opzioni del repository di entità](#) a pag. 18.

Tag sorgente. Un elenco di tag che indica le sorgenti dati attualmente note al repository. Scegliere una voce dall'elenco oppure <Aggiungi nuovo tag sorgente...> per creare un tag per una nuova sorgente dati.

Tipo di entità. Un elenco di tipi di entità (ovvero insiemi di funzionalità) definiti nel repository. Scegliere un tipo dall'elenco oppure <Aggiungi nuovo tipo di entità...> per visualizzare la finestra di configurazione del repository, in cui è possibile definire un nuovo tipo di entità. Per ulteriori informazioni, vedere l'argomento [Configurazione di un repository di entità](#) a pag. 22.

Chiave univoca. (obbligatorio) Il campo di input da utilizzare per gli identificatori univoci per i record di dati.

Tabella Mappatura. In questa tabella è possibile mappare ogni campo di input a una funzione corrispondente nel repository. Se non esiste una funzione adeguata nel tipo di entità selezionato, è possibile crearne qui una nuova.

- **Campo.** L'insieme dei campi di input nella sorgente dati selezionata. Ogni campo ha un'icona che indica il livello di misurazione (ovvero il tipo di dati) per il campo.
- **Mappato a funzione.** Per mappare un campo a una funzione, fare doppio clic su questa colonna (o premere la barra spaziatrice) nella riga del campo e scegliere una funzione dall'elenco. Se non è disponibile una funzione adeguata, scegliere <Aggiungi nuova funzione...> per visualizzare la finestra di configurazione del repository in cui è possibile definire una nuova funzione per questo tipo di entità. Per ulteriori informazioni, vedere l'argomento [Configurazione di un repository di entità](#) a pag. 22.
- **Utilizzo.** Indica il contesto di un campo particolare quando sono possibili più contesti, ad esempio i numeri di telefono di casa e ufficio. Esistono tipi di utilizzo preimpostati per le funzioni ADDRESS e PHONE; inoltre, si possono creare dei tipi di utilizzo personali per tutte le funzioni. Per impostare un utilizzo diverso da quello di default (Auto), fare clic su questa colonna nella riga desiderata e scegliere uno dei tipi di utilizzo esistenti (se disponibili), oppure fare clic su [Aggiungi utilizzo...](#) per crearne uno nuovo. Per ulteriori informazioni, vedere l'argomento [Gestione dei tipi di entità](#) a pag. 28.

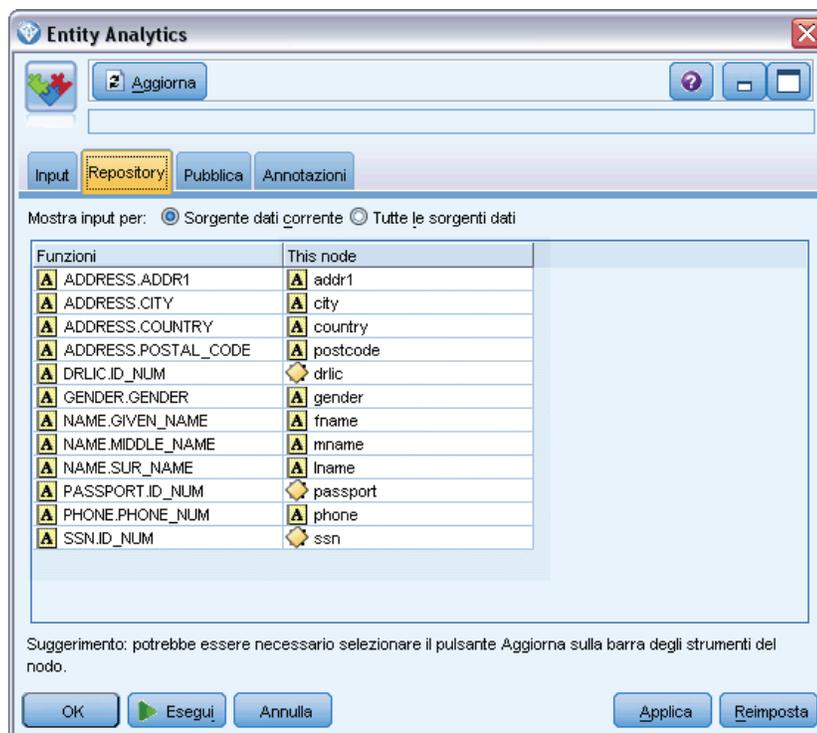
Importa mappatura. Importa un insieme precedentemente esportato di mappature campo-funzione da un file XML esterno. Questa operazione è utile se diverse sorgenti dati hanno gli stessi requisiti di mappatura, poiché evita di definire nuovamente le stesse mappature per diverse sorgenti.

Esporta mappatura. Esporta in un file XML esterno l'insieme di mappature campo-funzione indicate nella tabella di mappatura.

Visualizzazione delle mappature dei campi (nodo Esportazione EA)

Nella scheda Repository, fare clic sul pulsante [Aggiorna](#) per vedere per quali funzioni del repository sono stati mappati dei campi di input. Questa informazione può essere visualizzata per la sorgente dati corrente (quella controllata dal nodo di input collegato a questo nodo di esportazione) o per tutte le sorgenti dati.

Figura 3-4
Visualizzazione delle mappature dei campi



Mostra input per. Scegliere un'opzione per visualizzare le mappature per la sorgente dati corrente o per tutte le sorgenti dati note al repository.

Aggiorna. Aggiorna la visualizzazione per l'opzione di input selezionata.

Caratteristiche. Elenco di tutte le funzioni con mappature nelle sorgenti dati visualizzate. Le funzioni non mappate non vengono visualizzate.

<Sorgente dati>. In ogni colonna sono indicati i campi mappati in una determinata sorgente dati per ogni funzione per cui è stata definita una mappatura.

Configurazione di un repository di entità

È possibile gestire il contenuto del repository dalla finestra Configurazione repository, che fornisce un'interfaccia grafica di facile uso per l'intero repository.

se si utilizzano più repository con configurazioni uguali o simili, è possibile impostare una configurazione di base ed esportarla in un file che in seguito potrà essere importato in altri repository. Per ulteriori informazioni, vedere l'argomento [Riutilizzo di una configurazione di repository](#) a pag. 31.

Attenzione: Se si modifica e salva la configurazione di un repository che contiene dei dati, è assolutamente consigliabile eliminare il contenuto del repository e ricaricare i dati. In questo modo si evita di lasciare il repository in uno stato incoerente.

Per impostare una configurazione di repository

1. Aprire un nodo Entity Analytics.
2. Fare clic sull'elenco Repository entità.
3. Fare clic su <Cerca...> per visualizzare la finestra di dialogo Istanze di risoluzione entità.
4. Nella finestra di dialogo Istanze di risoluzione entità, fare clic sull'elenco Nome repository.
5. Selezionare il repository per cui si desidera impostare la configurazione.
6. Se non è già stata eseguita la connessione, immettere nome utente e password dell'amministratore e fare clic su Connetti.
7. Quando il pulsante Configura repository è attivato, fare clic su di esso per visualizzare la finestra Configurazione repository.
8. Creare i dettagli della configurazione come indicato nelle sezioni successive.

Il riquadro di spostamento a sinistra nella finestra Configurazione repository contiene una struttura ad albero da cui è possibile gestire le diverse caratteristiche del repository.

Tabella 3-1
Elementi principali della finestra Configurazione repository

Sezione	Descrizione	
Sorgenti dati	Visualizza le mappature di tutte le sorgenti dati alle varie funzioni del repository.	Per ulteriori informazioni, vedere l'argomento Visualizzazione delle mappature delle sorgenti dati a pag. 23.
Funzioni	Consente di creare una nuova funzione o di duplicare, modificare o eliminare una funzione già esistente.	Per ulteriori informazioni, vedere l'argomento Gestione delle funzioni del repository a pag. 24.
Tipi di entità	Consente di creare un nuovo tipo di entità o di gestire quelli già esistenti (duplicazione, ridenominazione, collegamento o rimozione di funzioni, eliminazione).	Per ulteriori informazioni, vedere l'argomento Gestione dei tipi di entità a pag. 28.
Regole di risoluzione	Consente di impostare la soglia per la corrispondenza delle entità.	Per ulteriori informazioni, vedere l'argomento Impostazione della soglia per la corrispondenza delle entità a pag. 30.

Visualizzazione delle mappature delle sorgenti dati

Nella sezione Sorgenti dati della finestra Configurazione repository, la voce Tutte le sorgenti fornisce una visualizzazione di sola lettura di tutte le mappature di tutte le sorgenti dati alle varie funzioni del repository.

Figura 3-5
Mappature delle sorgenti dati nella finestra di configurazione



Fare clic su **Aggiorna** per aggiornare l'elenco nel caso in cui siano state aggiunte nuove sorgenti dati al repository.

Nota: non è possibile aggiungere qui una sorgente dati al repository. Le sorgenti dati possono essere aggiunte solo creando un nodo di input di SPSS Modeler e collegandolo a un nodo di esportazione di Entity Analytics. Per ulteriori informazioni, vedere l'argomento [Connessione a una sorgente dati](#) a pag. 15.

Gestione delle funzioni del repository

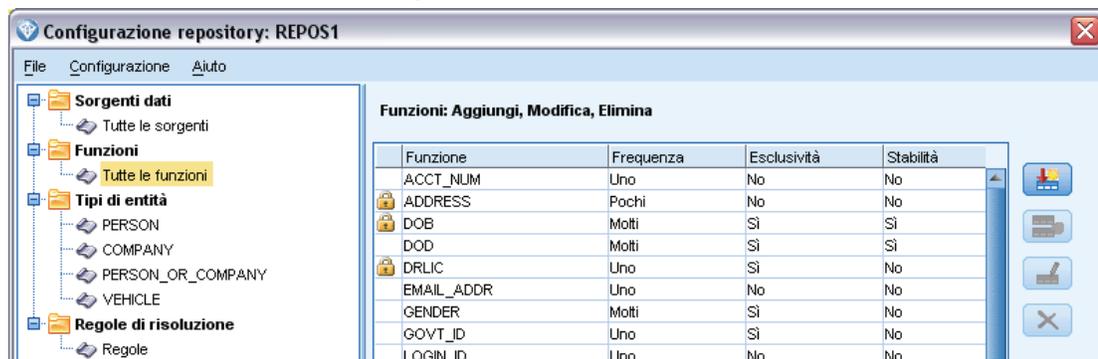
Una funzione di repository è un singolo tipo di informazione che può essere utilizzato con una sorgente dati di entità. Alcune funzioni (per esempio, Nome, Cognome, Data di nascita e così via) possono essere utilizzate da molte sorgenti dati diverse, mentre altre funzioni sono specifiche di una sorgente dati particolare. Una funzione può contenere uno o più elementi; normalmente ogni elemento equivale a un campo di un record di dati o a una colonna in una tabella di database.

Nella sezione Funzioni della finestra Configurazione repository, la voce Tutte le funzioni consente di gestire tutte le funzioni del repository. Possono essere eseguite le seguenti operazioni.

- Creare una nuova funzione
- Duplicare una funzione già esistente (per esempio, per creare una nuova funzione basata su una funzione esistente)
- Modificare una funzione esistente
- Eliminare una funzione esistente

Le istruzioni per queste operazioni verranno fornite più avanti in questa sezione.

Figura 3-6
Funzioni indicate nella finestra di configurazione



Nell'elenco delle funzioni sono riportate tutte le funzioni definite nel repository. Le colonne nell'elenco mostrano le diverse proprietà di cui può disporre una funzione.

Funzione. Il nome della funzione. Il simbolo del lucchetto accanto al nome di una funzione indica che la funzione è bloccata. Le funzioni bloccate non possono essere eliminate, duplicate né salvate con modifiche.

Frequenza. Indica quante identità possono avere lo stesso valore per questa funzione. I valori validi sono Nessuno, Nome, Uno (ad esempio, per un numero di passaporto), Pochi (ad esempio, per un indirizzo), oppure Molti (ad esempio, per una data di nascita).

Esclusività. Indica che un'entità dovrebbe avere, in genere, solo un tipo di questa funzione. Ad esempio, una data di nascita o un codice fiscale avrebbe qui il valore Si, mentre un indirizzo o carta di credito avrebbe il valore No (in quanto un'entità può presentare più indirizzi o carte di credito).

Stabilità. Indica il valore di stabilità di questa funzione (vale a dire, se è *improbabile* che cambi durante la durata dell'entità). Ad esempio, una funzione data di nascita avrebbe il valore Si perché non cambia mai, ma un indirizzo avrebbe il valore No perché è abbastanza probabile che vari nel tempo, il che lo rende meno stabile. *Nota:* Il sesso, inteso come genere, è generalmente stabile ma poiché viene spesso specificato in modo errato da dati non precisi, la configurazione di default propone il valore No.

Per creare una nuova funzione

1. Effettuare una delle seguenti operazioni.
 - Fare clic sul pulsante Crea nuova funzione (il pulsante in alto a destra sullo schermo).
 - Fare clic con il pulsante destro del mouse su Tutte le funzioni nel riquadro di spostamento a sinistra e scegliere Nuova funzione.
2. Completare la finestra di dialogo Aggiungi/Modifica funzione. Per ulteriori informazioni, vedere l'argomento [Aggiunta o modifica di una funzione](#) a pag. 26.

Per duplicare una funzione esistente

1. Nella colonna Funzione della tabella a destra sullo schermo, selezionare la funzione da duplicare.

2. Fare clic sul pulsante Duplica funzione selezionata (il secondo pulsante a destra sullo schermo).
3. Completare la finestra di dialogo Aggiungi/Modifica funzione. Per ulteriori informazioni, vedere l'argomento [Aggiunta o modifica di una funzione](#) a pag. 26.

Per modificare una funzione esistente

Attenzione: se si modifica o elimina una funzione o l'elemento di una funzione quando il repository contiene dei dati, sarà necessario successivamente svuotare il repository e ricaricare i dati. In questo modo si evita di lasciare il repository in uno stato incoerente.

1. Nella colonna Funzione della tabella a destra sullo schermo, selezionare la funzione da modificare. *Nota:* è possibile modificare solo le funzioni create, non quelle fornite dal sistema.
2. Fare clic sul pulsante Modifica funzione selezionata (il terzo pulsante a destra sullo schermo).
3. Completare la finestra di dialogo Aggiungi/Modifica funzione. Per ulteriori informazioni, vedere l'argomento [Aggiunta o modifica di una funzione](#) a pag. 26.

Per eliminare una funzione esistente

Attenzione: se si modifica o elimina una funzione o l'elemento di una funzione quando il repository contiene dei dati, sarà necessario successivamente svuotare il repository e ricaricare i dati. In questo modo si evita di lasciare il repository in uno stato incoerente.

1. Nella colonna Funzione della tabella a destra sullo schermo, selezionare la funzione da eliminare. *Nota:* è possibile eliminare solo le funzioni create, non quelle fornite dal sistema.
2. Effettuare una delle seguenti operazioni.
 - Fare clic sul pulsante Elimina funzione selezionata (l'ultimo pulsante in basso a destra sullo schermo).
 - Fare clic con il pulsante destro del mouse su Tutte le funzioni nel riquadro di spostamento a sinistra e scegliere Elimina.
3. Fare clic su Continua per confermare l'eliminazione della funzione. *Nota:* non è possibile annullare l'eliminazione di una funzione.

Aggiunta o modifica di una funzione

Attenzione: se si modifica o elimina una funzione o l'elemento di una funzione quando il repository contiene dei dati, sarà necessario successivamente svuotare il repository e ricaricare i dati. In questo modo si evita di lasciare il repository in uno stato incoerente.

Nella finestra di dialogo Aggiungi/modifica funzione è possibile creare una nuova funzione del repository oppure duplicare o modificare una funzione già esistente.

Figura 3-7
Modifica di una funzione

Aggiungi/Modifica funzione

Tipo di funzione: ADDRESS

Descrizione: Address

Frequenza: Pochi Esclusività: No

Stabilità nel tempo: No Mantieni cronologia: Sì

	Elemento	Descrizione	Tipo di dati	Cerca	Confronta	Descrittivo
1	ADDR_TYPE	Address t...	Stringa	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
2	FULL_ADDR	Full address	Stringa	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
3	ADDR1	Addr line 1	Stringa	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
4	ADDR2	Addr line 2	Stringa	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
5	ADDR3	Addr line 3	Stringa	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
6	CITY	City	Stringa	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
7	STATE	State	Stringa	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
8	POSTAL_...	Postal code	Stringa	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
9	COUNTRY	Country	Stringa	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>

OK Annulla Ajuto

Tipo di funzione. Etichetta che indica il tipo di informazioni a cui fa riferimento la funzione. Questa etichetta costituisce la prima parte dell'identificatore della funzione.

Descrizione. Breve descrizione del tipo di funzione, solo a scopo informativo.

Frequenza. Indica quante identità possono avere lo stesso valore per questa funzione. I valori validi sono Nessuno, Nome, Uno (ad esempio, per un numero di passaporto), Pochi (ad esempio, per un indirizzo), oppure Molti (ad esempio, per una data di nascita).

Esclusività. Indica che un'entità dovrebbe avere, in genere, solo un tipo di questa funzione. Ad esempio, una data di nascita o un codice fiscale avrebbe qui il valore Sì, mentre un indirizzo o carta di credito avrebbe il valore No (in quanto un'entità può presentare più indirizzi o carte di credito).

Stabilità. Indica il valore di stabilità di questa funzione (vale a dire, se è *improbabile* che cambi durante la durata dell'entità). Ad esempio, una funzione data di nascita avrebbe il valore Sì perché non cambia mai, ma un indirizzo avrebbe il valore No perché è abbastanza probabile che vari nel tempo, il che lo rende meno stabile. *Nota:* Il sesso, inteso come genere, è generalmente stabile ma poiché viene spesso specificato in modo errato da dati non precisi, la configurazione di default propone il valore No.

Mantieni cronologia. Se impostata su Sì (default), mantiene la cronologia delle modifiche apportate al valore del campo. Questa opzione risulta utile, ad esempio, per il campo indirizzo di un cliente. In questo caso, mantenendo la cronologia delle modifiche sarà più facile trovare le corrispondenze se il cliente ha cambiato indirizzo più volte. Tuttavia, per un campo tipo il saldo di un conto, è preferibile impostare l'opzione su No per evitare di conservare quantità esagerate di dati.

Tabella degli elementi. Elenco degli elementi compresi nella funzione.

■ **Elemento.** Nome dell'elemento.

- **Descrizione.** Breve descrizione delle opzioni offerte dall'elemento.
- **Tipo di dati.** Il tipo di dati che possono essere utilizzati per questo elemento.
- **Cerca.** Quando questa opzione è selezionata, significa che questo elemento della funzione può essere cercato da solo o in combinazione con altri elementi della ricerca di funzioni. Ad esempio, supponiamo che una funzione PASSPORT contenga due elementi: ID_NUM e COUNTRY. Per poter cercare l'elemento ID_NUM da solo, si selezionerà la sua casella Cerca lasciando vuota la casella Cerca di COUNTRY. Tutti i modificatori della funzione, quale paese o stato, o data di emissione e di scadenza, non verranno inclusi nella ricerca.
- **Confronta.** Indica quali elementi della funzione verranno inviati alla routine di confronto per la risoluzione delle entità. Rimanendo sull'esempio della funzione PASSPORT con due elementi, ID_NUM e COUNTRY, poiché COUNTRY potrebbe non essere sempre presente, solo ID_NUM deve essere inviato alla routine di confronto.
- **Descrittivo.** Indica quali elementi della funzione devono apparire nella sua descrizione interna. Nell'esempio precedente dell'ipotetica funzione PASSPORT con due elementi, ID_NUM e COUNTRY, per far apparire entrambi nella descrizione della funzione, si dovrebbe selezionare la casella Descrittivo di entrambi. Supponiamo, però, che la funzione presenti l'elemento aggiuntivo ISSUE_DT. La casella Descrittivo di quest'ultimo dovrebbe rimanere deselezionata per fare in modo che l'elemento non compaia nella descrizione della funzione.

Aggiungi nuovo elemento. Aggiunge una nuova riga alla tabella degli elementi per consentire la definizione di un nuovo elemento.

Elimina elemento. Elimina una riga selezionata dalla tabella degli elementi. Questa operazione non può essere annullata.

Attenzione: se si modifica o elimina una funzione o l'elemento di una funzione quando il repository contiene dei dati, sarà necessario successivamente svuotare il repository e ricaricare i dati. In questo modo si evita di lasciare il repository in uno stato incoerente.

Gestione dei tipi di entità

Un **tipo di entità** è un insieme denominato di funzioni del repository che possono essere riunite logicamente. Per esempio, un tipo di entità da utilizzare per l'insieme di dati di un cliente può essere costituito da funzioni come Nome, Data di nascita, Sesso, Indirizzo, Numero di telefono e così via.

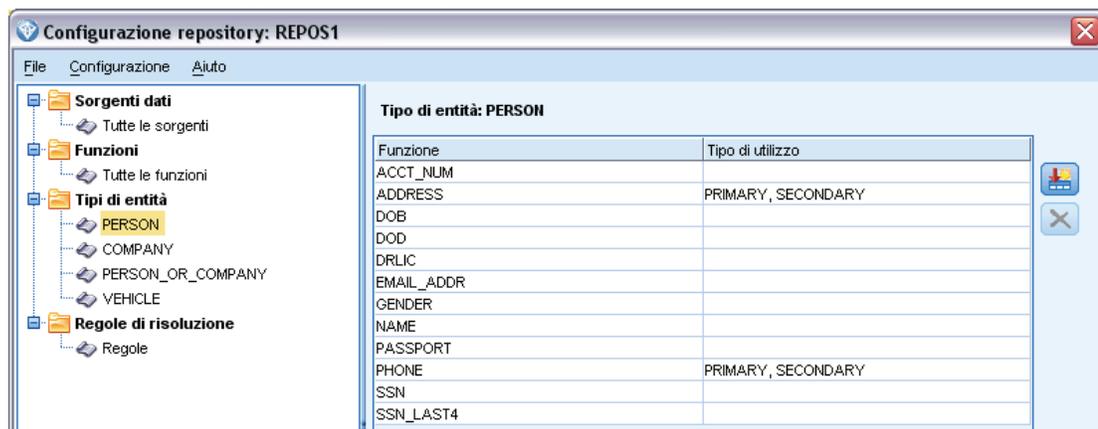
Il repository IBM SPSS Modeler Entity Analytics viene fornito con un insieme standard di tipi di entità, a cui se ne possono aggiungere altri.

La sezione Tipi di entità della finestra Configurazione repository contiene i diversi tipi di entità creati. Possono essere eseguite le seguenti operazioni.

- Creare un nuovo tipo di entità
- Duplicare un tipo di entità già esistente (per esempio, per creare un nuovo tipo di entità basato su un tipo esistente)
- Collegare funzioni a un tipo di entità
- Rimuovere funzioni da un tipo di entità

- Rinominare un tipo di entità
- Eliminare un tipo di entità

Figura 3-8
Elenco delle funzioni appartenenti a un tipo di entità



Tipo di entità. Il nome del tipo di entità selezionato.

Funzione. L'elenco delle funzioni valide comprese in questo tipo di entità.

Tipo di utilizzo. (Facoltativo) Indica diversi contesti in cui è possibile utilizzare questa funzione. Fare doppio clic su questa colonna per aggiungere o modificare un tipo di utilizzo, separando i tipi di utilizzo con una virgola e uno spazio. Il valore specificato in questo punto determina i valori visualizzati nei nodi Esportazione EA o Streaming di EA quando un utente fa clic sulla colonna Utilizzo per una funzione nella scheda Input. Per ulteriori informazioni, vedere l'argomento [Opzioni di input del repository per la mappatura](#) a pag. 20.

Per creare un nuovo tipo di entità

1. Fare clic con il pulsante destro del mouse su Tipi di entità nel riquadro di spostamento visualizzato a sinistra.
2. Scegliere Nuovo tipo di entità.
3. Inserire un nome univoco per il tipo di entità e fare clic su OK.
4. Collegare le funzioni al tipo di entità (vedere la sezione successiva).

Per collegare funzioni a un tipo di entità

1. Selezionare il tipo di entità nel riquadro di spostamento visualizzato a sinistra.
2. Fare clic sul pulsante Collega funzione (il pulsante in alto a destra sullo schermo).
3. Scegliere una o più opzioni dall'elenco delle funzioni disponibili (utilizzare Ctrl-clic per scegliere più funzioni) e fare clic su OK.

Per rimuovere le funzioni da un tipo di entità

1. Selezionare il tipo di entità nel riquadro di spostamento visualizzato a sinistra.
2. Selezionare una o più funzioni dalla tabella delle funzioni collegate visualizzata a destra. Utilizzare Ctrl-clic per scegliere più funzioni.
3. Fare clic sul pulsante Scollega funzione (il pulsante in basso a destra sullo schermo).

Per duplicare un tipo di entità esistente

1. Nel riquadro di spostamento a sinistra, fare clic con il pulsante destro del mouse sul tipo di entità da duplicare.
2. Scegliere Duplica tipo di entità.
3. Inserire un nome univoco per il nuovo tipo di entità e fare clic su OK.
4. Collegare o rimuovere le funzioni al o dal tipo di entità secondo le proprie esigenze (vedere le istruzioni precedenti).

Per rinominare un tipo di entità

Attenzione: se si modifica o elimina una funzione o l'elemento di una funzione quando il repository contiene dei dati, sarà necessario successivamente svuotare il repository e ricaricare i dati. In questo modo si evita di lasciare il repository in uno stato incoerente.

1. Nel riquadro di spostamento a sinistra, fare clic con il pulsante destro del mouse sul tipo di entità da rinominare.
2. Scegliere Rinomina.
3. Inserire il nuovo nome per il tipo di entità e fare clic su OK.

Per eliminare un tipo di entità

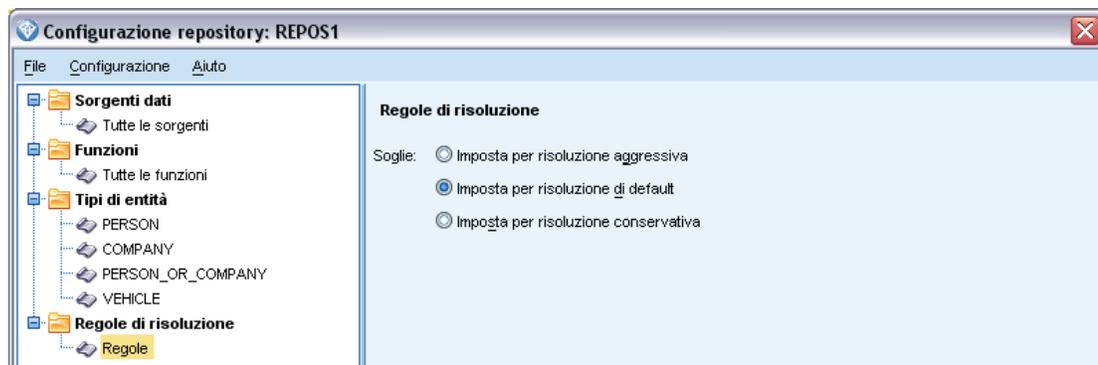
Attenzione: se si modifica o elimina una funzione o l'elemento di una funzione quando il repository contiene dei dati, sarà necessario successivamente svuotare il repository e ricaricare i dati. In questo modo si evita di lasciare il repository in uno stato incoerente.

1. Nel riquadro di spostamento a sinistra, fare clic con il pulsante destro del mouse sul tipo di entità da eliminare.
2. Scegliere Elimina.
3. Fare clic su OK per confermare l'eliminazione del tipo di entità. *Attenzione:* non è possibile annullare l'eliminazione di un tipo di entità.

Impostazione della soglia per la corrispondenza delle entità

Nella sezione Regole di risoluzione della finestra Configurazione repository, è possibile scegliere la soglia in base alla quale si ottiene la corrispondenza delle entità.

Figura 3-9
Impostazione delle regole di risoluzione per le entità



Quando si crea il repository, la corrispondenza viene preimpostata sulla soglia di default.

Scegliere Imposta per risoluzione aggressiva se non si trovano nei record corrispondenze sufficienti per eseguire la risoluzione delle entità.

Scegliere Imposta per risoluzione di default per tornare alla soglia di default da una delle altre impostazioni.

Scegliere Imposta per risoluzione conservativa se vengono trovate troppe corrispondenze.

Riutilizzo di una configurazione di repository

Se è già stata impostata una configurazione e si desidera utilizzarla per un altro repository, è possibile esportare la configurazione esistente in un file XML e importare il file nell'altro repository (destinazione).

Per riutilizzare una configurazione esistente

1. Visualizzare la finestra Configurazione repository per il repository di cui si desidera utilizzare la configurazione. Per ulteriori informazioni, vedere l'argomento [Configurazione di un repository di entità](#) a pag. 22.
2. Dal menu visualizzato nella finestra scegliere Configurazione > Esporta configurazione.
3. Nella finestra di dialogo Salva con nome, scegliere il nome e il percorso del file XML di esportazione.
4. Visualizzare la finestra Configurazione repository per il repository di destinazione.
5. Dal menu visualizzato nella finestra scegliere Configurazione > Importa configurazione.
6. Nella finestra di dialogo Apri, scegliere il nome e il percorso del file XML esportato in precedenza e fare clic su Apri.

Salvataggio delle modifiche alla configurazione

Per salvare le modifiche alla configurazione

Dal menu della finestra Configurazione repository scegliere
File > Salva.

Chiusura della finestra di configurazione

Per uscire dalla finestra di configurazione

Dal menu della finestra Configurazione repository scegliere
File > Esci.

Se alcune modifiche apportate alla configurazione non sono state salvate, fare clic su OK per salvarle e uscire oppure su Annulla per uscire senza salvare.

Analisi delle identità risolte (nodo di input Entity Analytics(EA))

Avendo i dati di almeno una sorgente presenti nel repository, è possibile utilizzare il nodo di input Entity Analytics(EA) per passare le identità risolte ad altri nodi di IBM® SPSS® Modeler per ulteriori analisi o elaborazioni, per esempio la creazione di un report che elenca le identità risolte.

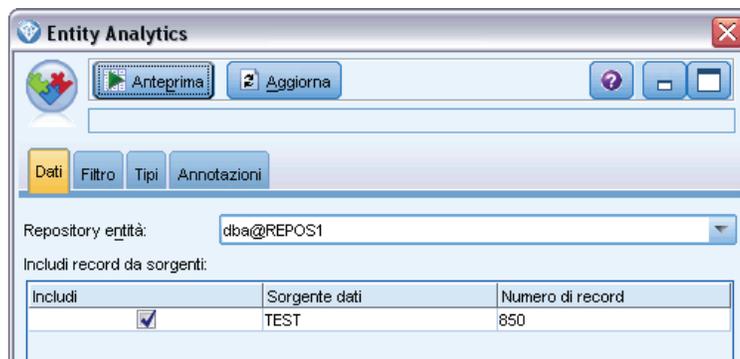
Per analizzare le identità risolte

1. Aggiungere un nodo di input Entity Analytics(EA) a uno stream.
2. Aprire il nodo Entity Analytics(EA).
3. Nella scheda Dati, selezionare il repository di entità e una o più delle relative sorgenti dati di input (fare clic su **Aggiorna** per aggiornare il conteggio dei record). Per ulteriori informazioni, vedere l'argomento [Selezione di una sorgente dati](#) a pag. 32.
4. Aggiungere ulteriori nodi allo stream per eseguire l'elaborazione desiderata. Per ulteriori informazioni, vedere l'argomento [Aggiunta di nodi allo stream](#) a pag. 34.

Selezione di una sorgente dati

Nella scheda Dati, deve essere selezionata almeno una sorgente dati del repository su cui verranno eseguite ulteriori elaborazioni. Fare clic su **Aggiorna** per aggiornare i conteggi dei record per le sorgenti dati elencate.

Figura 3-10
Selezione di una sorgente dati nel nodo di input



Repository entità. Visualizza il repository di entità corrente, se esiste. Per scegliere un repository diverso nel caso in cui ne esista più di uno, selezionarlo dall'elenco. Per creare un nuovo repository, scegliere <Cerca...> per visualizzare una finestra di dialogo da cui è possibile creare il repository. Per ulteriori informazioni, vedere l'argomento [Opzioni del repository di entità](#) a pag. 18.

Includi record da sorgenti. In questa tabella sono elencate le diverse sorgenti dati che sono state inserite nel repository, unitamente al numero di record in ogni sorgente. Selezionare la casella di controllo Includi per le sorgenti dati che si desidera utilizzare per ulteriori analisi ed elaborazioni.

Ridenominazione dei campi dati

È possibile utilizzare la scheda Filtro per rinominare qualsiasi campo delle identità risolte passato a valle per un'ulteriore elaborazione. È possibile rinominare un campo di identità risolta, ad esempio, per mantenere la compatibilità dei nomi dei campi quando si effettua un'unione a valle con un altro insieme di dati.

I campi con i rispettivi nomi originali sono riportati di seguito.

Tabella 3-2
Campi identità risolte

Campo	Descrizione
\$EA-ID	Identificatore entità
\$EA-SRC	Tag sorgente che identifica la sorgente dati nella quale hanno avuto origine i record
\$EA-KEY	Campo designato come chiave univoca nel file della sorgente dati

Nota: benché sia possibile utilizzare anche la scheda Filtri per eliminare i campi, non è possibile farlo in questo caso, dal momento che i campi delle identità risolte sono il minimo necessario per l'elaborazione dell'analisi delle entità.

Impostazione delle informazioni sui tipi per i campi dati

Nella scheda Tipi, è possibile visualizzare o modificare diverse proprietà dei campi delle identità risolte passati a valle per un'ulteriore elaborazione.

Le proprietà che è possibile modificare sono le stesse della scheda Tipi di un nodo Tipo di SPSS Modeler normale e sono riportate di seguito.

Tabella 3-3
Proprietà dei tipi per i campi

Proprietà	Descrizione
Misurazione	Il livello di misurazione (cioè il tipo di dati) utilizzato per descrivere le caratteristiche dei dati nel campo.
Valori	Fornisce le opzioni per leggere i valori dei dati dall'insieme di dati.
Mancante	Utilizzata per specificare come vengono gestiti i valori mancanti del campo.
Controllo	Opzioni di convalida per assicurare che i valori dei campi siano conformi ai valori o agli intervalli specificati.
Ruolo	Specifica come verrà utilizzato il campo se i dati vengono passati a un nodo Modelli o a un insieme di modelli.

Aggiunta di nodi allo stream

È possibile aggiungere vari nodi di SPSS Modeler allo stream per eseguire operazioni di analisi o elaborazione dell'output del nodo di input Entity Analytics(EA). Ad esempio, è possibile aggiungere uno o più degli elementi seguenti.

- Nodo Aggregazione per riassumere l'output, che può essere di dimensioni molto grandi
- Nodo Seleziona per selezionare un sottoinsieme dell'output
- Nodo Tabella per visualizzare l'output del nodo di input Entity Analytics(EA)
- Nodo Report per stampare l'output in un report
- Nodo di esportazione di SPSS Modeler per esportare l'output in un formato differente, per esempio un foglio di calcolo o un database

Per ulteriori informazioni, vedere le sezioni relative ai nodi di operazioni su record, output ed esportazione nei *Manuali relativi ai nodi di Input, Elaborazione e Output di IBM SPSS Modeler*.

Confronto di nuovi casi con il repository (nodo Streaming di EA)

Una volta effettuata la risoluzione delle identità nel repository, è possibile utilizzare il nodo Streaming di EA per confrontare nuovi casi incontrati successivamente con il contenuto del repository. Questo nodo elabora i record provenienti da una nuova sorgente dati, li confronta con le entità risolte già presenti nel repository e trasmette i record per cui esistono delle corrispondenze per un'ulteriore elaborazione. Le corrispondenze rispetto alle entità esistenti possono essere esatte oppure parziali.

Come il nodo Esportazione EA, il nodo Streaming di EA prende un singolo nodo di input di SPSS Modeler come input. Tuttavia, il nodo Streaming di EA presenta le differenze riportate di seguito. Mentre il nodo di esportazione genera record per tutte le entità correlate ai rispettivi record di

input, il nodo Streaming di EA genera record soltanto per le entità correlate a entità già risolte nel repository. Per ulteriori informazioni, vedere l'argomento [Output del nodo Streaming di EA](#) a pag. 39.

Per confrontare i nuovi casi in base al repository

1. Connettersi alla sorgente dati contenente i nuovi record da confrontare con le entità esistenti. Per ulteriori informazioni, vedere l'argomento [Connessione a una sorgente dati](#) a pag. 15.
2. Dalla scheda Oper su record, collegare un nodo Streaming di EA al nodo di input dei dati.
3. Fare doppio clic sul nodo di esportazione di Entity Analytics per aprire la relativa finestra di dialogo.
4. Fare clic sull'elenco Repository entità.
5. Fare clic su <Cerca...> per visualizzare la finestra di dialogo Repository entità.
6. Nella finestra di dialogo Repository entità, fare clic sul campo Nome repository.
7. Fare clic sul nome del repository da usare.
8. Immettere il nome utente e la password di tale repository e fare clic su Connetti. Fare clic su OK quando il repository è connesso.
9. Nella finestra di dialogo Streaming di EA, selezionare il Tipo di entità da utilizzare. Per ulteriori informazioni, vedere l'argomento [Gestione dei tipi di entità](#) a pag. 28.
10. Mappare i campi di input nella sorgente dati alle funzioni del repository. Per ulteriori informazioni, vedere l'argomento [Mappatura dei campi di input alle funzioni \(nodo Streaming di EA\)](#) a pag. 36.
11. Se si desidera, fare clic sulla scheda Repository per visualizzare una tabella che mostra la mappatura di questa sorgente dati (oppure tutte le sorgenti dati che sono state mappate). Per ulteriori informazioni, vedere l'argomento [Visualizzazione delle mappature dei campi \(nodo Streaming di EA\)](#) a pag. 37.
12. Fare clic sulla scheda Visualizza per vedere i dettagli delle diverse sorgenti dati che sono state inserite nel repository e impostare il criterio di selezione per recuperare le entità esistenti. Per ulteriori informazioni, vedere l'argomento [Visualizzazione delle sorgenti dati \(nodo Streaming di EA\)](#) a pag. 38.
13. Fare clic su OK quando il nodo è impostato correttamente.
14. Collegare un nodo Tabella al nodo Streaming di EA ed eseguire lo stream.

La finestra dell'output del nodo Tabella elenca tutte le entità recuperate che corrispondono ai nuovi record nella sorgente dati. Ai campi di output viene aggiunto il prefisso \$EA-. Per ulteriori informazioni, vedere l'argomento [Output del nodo Streaming di EA](#) a pag. 39.

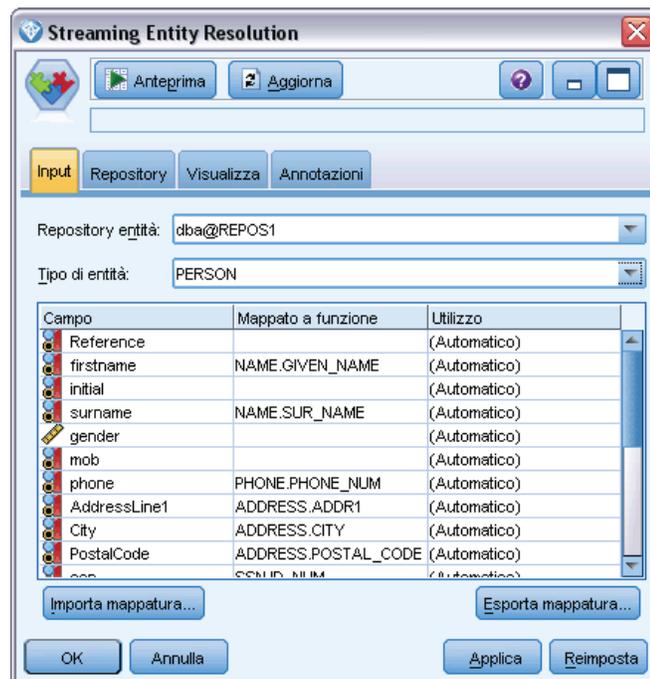
Mappatura dei campi di input alle funzioni (nodo Streaming di EA)

La scheda Input contiene le opzioni per la mappatura dei campi nell'input a questo nodo alle funzioni del repository. Impostare le assegnazioni della mappatura in questa scheda, oppure nella scheda Visualizza, per vedere i dettagli di tutte le sorgenti dati nel repository, quindi fare clic su OK.

Se si è già archiviato un insieme di mappature in un file XML, è possibile utilizzarle facendo clic su Importa mappatura.

Figura 3-11

Mappatura dei campi provenienti da nuovi record alle funzioni nel repository



Repository entità. Visualizza il repository di entità corrente, se esiste. Per scegliere un repository diverso nel caso in cui ne esista più di uno, selezionarlo dall'elenco. Per creare un nuovo repository, scegliere <Cerca...> per visualizzare una finestra di dialogo da cui è possibile creare il repository. Per ulteriori informazioni, vedere l'argomento [Opzioni del repository di entità](#) a pag. 18.

Tipo di entità. Un elenco di tipi di entità (ovvero insiemi di funzionalità) definiti nel repository. Scegliere un tipo dall'elenco oppure <Aggiungi nuovo tipo di entità...> per visualizzare la finestra di configurazione del repository, in cui è possibile definire un nuovo tipo di entità. Per ulteriori informazioni, vedere l'argomento [Configurazione di un repository di entità](#) a pag. 22.

Tabella Mappatura. In questa tabella è possibile mappare ogni campo di input a una funzione corrispondente nel repository. Se non esiste una funzione adeguata nel tipo di entità selezionato, è possibile crearne qui una nuova.

- **Campo.** L'insieme dei campi di input nella sorgente dati selezionata. Ogni campo ha un'icona che indica il livello di misurazione (ovvero il tipo di dati) per il campo.

- **Mappato a funzione.** Per mappare un campo a una funzione, fare doppio clic su questa colonna (o premere la barra spaziatrice) nella riga del campo e scegliere una funzione dall'elenco. Se non è disponibile una funzione adeguata, scegliere <Aggiungi nuova funzione...> per visualizzare la finestra di configurazione del repository in cui è possibile definire una nuova funzione per questo tipo di entità. Per ulteriori informazioni, vedere l'argomento [Configurazione di un repository di entità](#) a pag. 22.
- **Utilizzo.** Indica il contesto di un campo particolare quando sono possibili più contesti, ad esempio i numeri di telefono di casa e ufficio. Per ulteriori informazioni, vedere l'argomento [Gestione dei tipi di entità](#) a pag. 28.

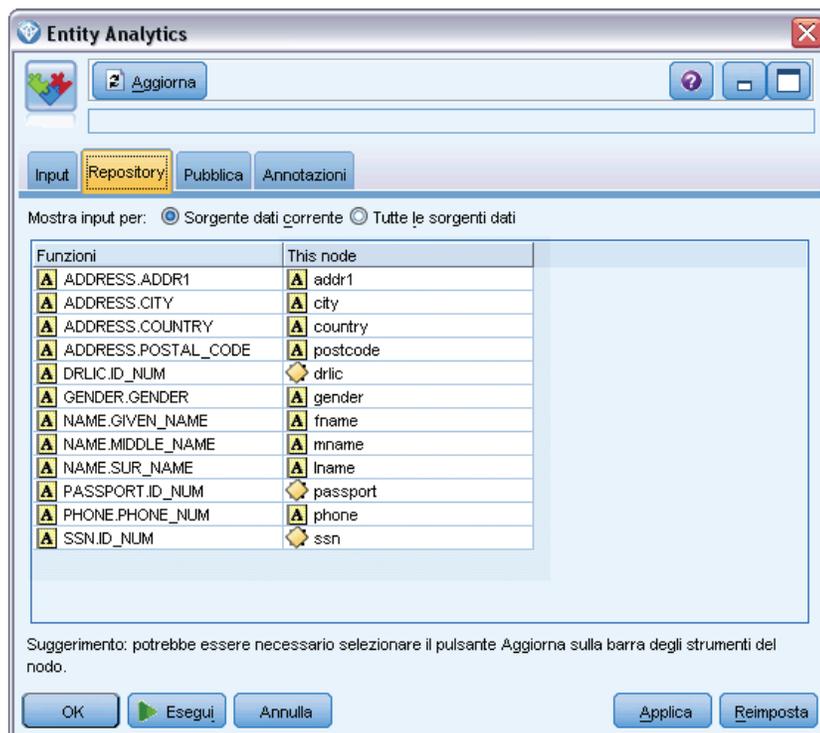
Importa mappatura. Importa un insieme precedentemente esportato di mappature campo-funzione da un file XML esterno. Questa operazione è utile se diverse sorgenti dati hanno gli stessi requisiti di mappatura, poiché evita di definire nuovamente le stesse mappature per diverse sorgenti.

Esporta mappatura. Esporta in un file XML esterno l'insieme di mappature campo-funzione indicate nella tabella di mappatura.

Visualizzazione delle mappature dei campi (nodo Streaming di EA)

Nella scheda Repository, fare clic su **Aggiorna** per vedere per quali funzioni del repository sono stati mappati dei campi di input. Questa informazione può essere visualizzata per la sorgente dati corrente (quella controllata dal nodo di input collegato a questo nodo di esportazione) o per tutte le sorgenti dati.

Figura 3-12
Visualizzazione delle mappature dei campi



Mostra input per. Scegliere un'opzione per visualizzare le mappature per la sorgente dati corrente o per tutte le sorgenti dati note al repository.

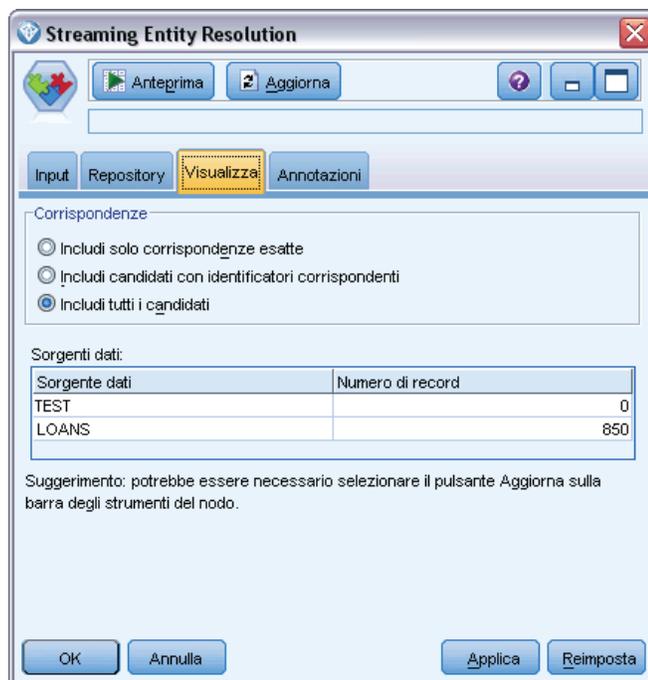
Caratteristiche. Elenco di tutte le funzioni con mappature nelle sorgenti dati visualizzate. Le funzioni non mappate non vengono visualizzate.

Questo nodo. In ogni colonna sono indicati i campi mappati in una determinata sorgente dati per ogni funzione per cui è stata definita una mappatura.

Visualizzazione delle sorgenti dati (nodo Streaming di EA)

Nella scheda Visualizza è possibile vedere i dettagli delle diverse sorgenti dati che sono state inserite nel repository. Si tratta delle sorgenti dati in base alle quali viene elaborato l'input a questo nodo, per cercare e recuperare le entità corrispondenti. Fare clic su Aggiorna per aggiornare i conteggi dei record.

Figura 3-13
Visualizzazione dei dettagli delle sorgenti dati



Corrispondenze. Queste opzioni specificano quanto le informazioni di mappatura campo-funzione specificate nella scheda Input devono corrispondere in base ai record candidati (cioè l'intero contenuto del repository). Più severo è il criterio di corrispondenza, minore sarà il numero di entità recuperate.

- **Includi solo corrispondenze esatte.** Si tratta del criterio di corrispondenza più severo e dà come risultato il minor numero di record selezionati. Utilizzare questa opzione per restituire solo le entità considerate corrispondenze esatte.

- **Includi candidati con identificatori corrispondenti.** Utilizzare questa impostazione per restituire le entità corrispondenti e le entità che condividono i medesimi identificatori (le cui funzioni sono state configurate con un valore di frequenza di uno, ad esempio, corrispondenze tra numeri di carte di credito, codici fiscali e così via).
- **Includi tutti i candidati.** Utilizzare questa opzione per vedere il numero di entità più ampio possibile del repository con funzioni condivise. Si tratta del criterio di corrispondenza meno severo e dà come risultato il maggior numero di record selezionati. Questa opzione restituisce corrispondenze esatte ed entità che condividono quasi tutte le funzioni (generalmente quelle con un valore di frequenza pari a Uno o Pochi). Ad esempio, verrebbero incluse sia entità con il medesimo codice fiscale che entità con indirizzi simili.

Sorgenti dati. In questa tabella sono elencati i tag sorgente delle diverse sorgenti dati che sono state caricate nel repository, unitamente al numero di record in ogni sorgente.

Output del nodo Streaming di EA

L'output del nodo Streaming di EA è costituito dai seguenti campi per ogni record recuperato.

Campo	Descrizione
<i>Campo1</i> [, <i>Campo2</i> [, ... <i>CampoN</i>]]	Campi della sorgente dati che contiene i nuovi record.
\$EA-ID	Identificatore di entità per questo record nel repository.
\$EA-SC	Campo che indica il livello di corrispondenza tra questo record e un'entità osservata nel repository; valore compreso tra 1.0 (corrispondenza scarsa) e 10.0 (corrispondenza buona).
\$EA-SRC	Tag sorgente che identifica la sorgente dati nella quale ha avuto origine il record.
\$EA-KEY	Valore di chiave univoca per questo record nel file della sorgente dati.
\$EA-Funzione1[, \$EA-Funzione2[, ... \$EA-FunzioneN]]	Valori delle funzioni mappate per questo record nel repository.

Utilizzo di IBM SPSS Modeler Entity Analytics con altri prodotti IBM SPSS

Sono disponibili programmi di installazione che consentono di utilizzare IBM SPSS Modeler Entity Analytics con i seguenti prodotti:

- IBM SPSS Collaboration and Deployment Services
- IBM SPSS Modeler Batch per Windows
- IBM SPSS Modeler Solution Publisher

È necessario aver eseguito i programmi di installazione prima di poter utilizzare le funzioni di IBM SPSS Modeler Entity Analytics con questi prodotti. Per ulteriori informazioni, fare riferimento al documento *IBM SPSS Modeler Premium - Guida all'installazione*.

Attività amministrative

Per i repository creati con Entity Analytics, viene creato un nuovo servizio database con il prodotto IBM solidDB. Vi sono alcune attività amministrative associate a solidDB, che in genere vengono eseguite dall'amministratore del database o dall'amministratore del sistema:

- Configurazione delle assegnazioni delle porte
- Gestione delle credenziali dell'amministratore per il database del repository

Le altre attività amministrative che possono essere necessarie riguardano tutti i repository e sono le seguenti:

- Spostamento della directory di archiviazione del repository
- Impostazione delle opzioni dello stream per i campi data/ora e timestamp
- Esecuzione di IBM SPSS Modeler Entity Analytics con il client SPSS Modeler e SPSS Modeler Server sul medesimo sistema Windows
- Svuotamento di un repository di entità
- Eliminazione di un repository di entità
- Eliminazione di un repository quando non è possibile connettersi a esso

Configurazione delle assegnazioni delle porte

È necessario che a ogni servizio database solidDB venga assegnata una porta che non può essere assegnata ad altri servizi in esecuzione sul computer. I servizi database risiedono nello stesso computer che esegue IBM® SPSS® Modeler Server (oppure, se IBM® SPSS® Modeler viene utilizzato senza una connessione a SPSS Modeler Server, nel computer che esegue IBM® SPSS® Modeler).

Per default, Entity Analytics assegna le porte nell'intervallo da 1320 a 1520, iniziando dalla porta 1320 per il primo repository creato. In caso di conflitto, è possibile configurare l'assegnazione delle porte modificando il file: *<percorso di installazione di modeler server>/ext/bin/pasw.entityanalytics/ea.cfg* e impostando i valori appropriati per le impostazioni *porta_min* e *porta_max*. Il contenuto di default di questo file è illustrato di seguito.

```
# port range configuration for entity analytics  
  
#  
  
# this port range controls which ports SolidDB databases  
  
# (created to store Entity Analytics Repositories in)  
  
# may use. Configure this if the default port range will  
  
# introduce a conflict on your system.  
  
#
```

```
# default min_port = 1320

# default max_port = 1520

min_port, 1320

max_port, 1520
```

Gestione delle credenziali dell'amministratore per il database del repository

Il nome utente e la password dell'amministratore del database solidDB che ospita un repository di entità vengono definiti quando viene creato il repository. Se si conoscono le credenziali correnti, è possibile modificare questi dettagli mediante l'editor SQL di solidDB.

Per avviare l'editor SQL di solidDB

1. Su un computer client, aprire una finestra del prompt dei comandi.
2. Immettere:

```
cd dir_install_modeler\ext\bin\pasw.entityanalytics\solidDB\bin
```

dove *dir_install_modeler* è la directory in cui è installato SPSS Modeler.

3. Immettere:

```
solsql -c "C:\Documents and Settings\All Users\Application  
Data\IBM\SPSS\Modeler\versione\EA\repositories\nome_repos
```

dove *versione* è il numero di versione dell'installazione SPSS Modeler e *nome_repos* è il nome del repository.

4. Al prompt, immettere il nome utente e la password correnti dell'amministratore del database per visualizzare il prompt `solsql>`.

Per modificare la password dell'amministratore del database

1. Al prompt `solsql>` immettere:

```
alter user nomeutente identified by password;  
commit work;
```

dove *nomeutente* è il nome utente corrente dell'amministratore del database e *password* è la nuova password.

2. Immettere Exit oppure chiudere l'editor.
3. Riavviare il client SPSS Modeler.

Per eseguire altre attività amministrative:

Per informazioni sulle altre attività amministrative da eseguire con il database solidDB, consultare la documentazione relativa alla versione appropriata di IBM solidDB all'indirizzo <http://publib.boulder.ibm.com/>.

Spostamento della directory di archiviazione del repository

Di default, i file del repository vengono archiviati nella directory *EA* nei percorsi seguenti:

- C:\Documents and Settings\All Users\ApplicationData\IBM\SPSS\Modeler\versione\EA (sistemi Windows)
- *dir_install_modeler/ext/bin/pasw.entityanalytics/EA* (sistemi UNIX)

Poiché i file utilizzati per archiviare il repository diventano molto grandi, può essere necessario spostarli su un disco o una partizione diversa per liberare dello spazio. A questo scopo, procedere come segue.

1. Uscire da SPSS Modeler.
2. Spostare la directory *EA* dalla sua posizione originale (indicata sopra) in una posizione diversa. Ad esempio, su Windows si potrebbe spostarla nel percorso *F:\data\EA*.
3. Modificare il file *<percorso di installazione di modeler server>/ext/bin/pasw.entityanalytics/ea.cfg* aggiungendo l'opzione seguente:

```
repository_data_directory, nuova_posizione
```

dove *nuova_posizione* rappresenta la directory in cui si è spostata la directory *EA*, ad esempio *F:\data\EA*.

Impostazione delle opzioni dello stream per i campi data/ora e timestamp

Se i dati sorgente contengono campi con valori data/ora o timestamp, verificare che le opzioni corrispondenti dello stream siano impostate sul formato riconosciuto da IBM SPSS Modeler Entity Analytics.

Per impostare il formato delle opzioni dello stream

1. Nel menu principale di SPSS Modeler, scegliere:
Strumenti > Proprietà stream > Opzioni.
2. Seleziona Data/ora.
3. Impostare Formato data su AAAA-MM-GG.
4. Impostare Formato ora su HH:MM:SS.
5. Fare clic su OK.

Esecuzione di IBM SPSS Modeler Entity Analytics con il client SPSS Modeler e SPSS Modeler Server sul medesimo sistema Windows

Se si è installato IBM SPSS Modeler Entity Analytics sia nel client SPSS Modeler che in SPSS Modeler Server sul medesimo sistema Windows, di default client e server condivideranno lo stesso repository. Per fare in modo che utilizzino repository separati, modificare il file di configurazione *ea.cfg* su **uno** dei sistemi configurandolo con un intervallo di porte e una cartella repository diversi.

Nota: è necessario seguire questa procedura in particolare se si usa il client SPSS Modeler a 32 bit e SPSS Modeler Server a 64 bit (o viceversa).

1. Aprire il file *<percorso di installazione di modeler [server]>/ext/bin/pasw.entityanalytics/ea.cfg* per modificarlo.
2. Modificare le impostazioni *min_port* e *max_port* su porte diverse rispetto all'altro sistema. Per ulteriori informazioni, vedere l'argomento [Configurazione delle assegnazioni delle porte](#) a pag. 40.
3. Modificare l'impostazione *repository_data_directory* su una directory diversa rispetto all'altro sistema.
4. Salvare e chiudere il file *ea.cfg*.

Svuotamento di un repository di entità

Se si desidera cancellare i record di dati da un repository di entità mantenendo le informazioni sulla configurazione, utilizzare l'opzione di svuotamento del repository.

Per svuotare un repository

1. Aprire un nodo Entity Analytics.
2. Fare clic sull'elenco Repository entità.
3. Fare clic su *<Cerca...>* per visualizzare la finestra di dialogo Istanze di risoluzione entità.
4. Nella finestra di dialogo Istanze di risoluzione entità, fare clic sull'elenco Nome repository.
5. Selezionare il repository che si desidera svuotare.
6. Se non è già stata eseguita la connessione, immettere nome utente e password dell'amministratore e fare clic su Connetti.
7. Quando il pulsante Svuota repository è abilitato, fare clic su di esso.
8. Fare clic su Svuota repository nella finestra di dialogo Svuota sorgenti dati per confermare lo svuotamento del repository.

Eliminazione di un repository di entità

Quando un repository non è più necessario, è possibile eliminarlo completamente.

Attenzione: l'eliminazione è definitiva. **Questa operazione non può essere annullata.** In caso di dubbio, utilizzare il pulsante *Svuota* per rimuovere tutti i dati di origine. In questo modo non verrà rimossa la configurazione del repository. Per ulteriori informazioni, vedere l'argomento [Svuotamento di un repository di entità](#) a pag. 43.

Nota: la seguente procedura presume che sia possibile collegarsi al repository da SPSS Modeler e che si conoscano il nome utente e la password dell'amministratore del database che ospita il repository. In caso contrario, attenersi alla procedura di eliminazione di un repository quando non è possibile connettersi a esso. Per ulteriori informazioni, vedere l'argomento [Eliminazione di un repository quando non è possibile connettersi a esso](#) a pag. 44.

Per eliminare un repository

1. Aprire un nodo Entity Analytics.
2. Fare clic sull'elenco Repository entità.
3. Fare clic su <Cerca...> per visualizzare la finestra di dialogo Istanze di risoluzione entità.
4. Nella finestra di dialogo Istanze di risoluzione entità, fare clic sull'elenco Nome repository.
5. Selezionare il repository che si desidera eliminare.
6. Se non è già stata eseguita la connessione, immettere nome utente e password dell'amministratore e fare clic su *Connetti*.
7. Quando il pulsante *Elimina intero repository* è abilitato, fare clic su di esso.
8. Fare clic su *Elimina* per confermare l'eliminazione del repository.
9. Fare clic *OK* per verificare che l'eliminazione sia stata eseguita.

Eliminazione di un repository quando non è possibile connettersi a esso

Attendersi alla seguente procedura se si desidera eliminare un repository di entità ma non è possibile connettersi a esso, per problemi di connettività con SPSS Modeler o perché si è dimenticato il nome utente o la password.

Eeguire la seguente procedura sul computer che ospita il database del repository.

Sistemi Windows

1. Aprire una finestra del prompt dei comandi.
2. Immettere:

```
cd dir_install_modeler  
cd ext\bin\pasw.entityanalytics\tools  
delete_repository.bat nome_repos
```

dove *dir_install_modeler* è la directory in cui è installato SPSS Modeler e *nome_repos* è il nome del repository.

3. Continuare da “Completamento della procedura” più avanti in questa sezione.

Sistemi UNIX

1. Aprire una shell.
2. Immettere:

```
cd dir_install_modeler_server  
cd ext/bin/pasw.entityanalytics/tools  
./delete_repository.sh nome_repos
```

dove *dir_install_modeler_server* è la directory in cui è installato SPSS Modeler Server e *nome_repos* è il nome del repository.

Completamento della procedura (tutti i sistemi)

1. Al prompt, confermare l’eliminazione del repository digitando Y.
2. Quando il repository è stato eliminato, viene visualizzato il messaggio:

Info - please remove repository files from the following directory:

percorso_directory

(Nota: potrebbe essere necessario riavviare il sistema prima di poter eliminare i file del repository)

3. Eliminare la directory con lo stesso nome del repository eliminato. Se non è possibile eliminare la directory, riavviare il computer e riprovare.

Funzionamento di Entity Analytics

Informazioni su questo esempio

In questo esempio vedremo come l'aggiunta dell'analisi delle entità possa ottimizzare ulteriormente i risultati straordinari offerti da IBM® SPSS® Modeler.

Questo esempio utilizza lo stream chiamato *loan_entity_analytics.str*, che fa riferimento al file di dati *loan_applications.csv*. Questi file sono disponibili nella directory *Demos* di tutte le installazioni di SPSS Modeler in cui è installato anche IBM® SPSS® Modeler Entity Analytics. È possibile accedere alla directory *Demos* dal gruppo di programmi SPSS Modeler nel menu Start di Windows. Il file *loan_entity_analytics.str* si trova nella directory *Entity_Analytics*.

Nota: Prima di poter eseguire questo stream di esempio, è necessario creare un repository sul proprio sistema. Provvedere subito prima di continuare con questo esempio. Per ulteriori informazioni, vedere l'argomento [Creazione del repository](#) in il capitolo 3 a pag. 16.

Lo scenario è una situazione familiare, in cui i dirigenti di una banca sono preoccupati per il fatto che probabilmente i clienti saranno inadempienti riguardo a prestiti per i quali sono ancora in corso le procedure di richiesta. Il reparto IT della banca utilizza da molto tempo SPSS Modeler, quindi ha già creato uno stream e un modello predittivo sulla base dei dati esistenti per circa 700 prestiti concessi in passato. Tali prestiti sono stati rimborsati oppure i clienti sono stati inadempienti.

Il modello originale

Ecco come il personale della banca ha creato il modello e quali indicazioni ha tratto da tale modello.

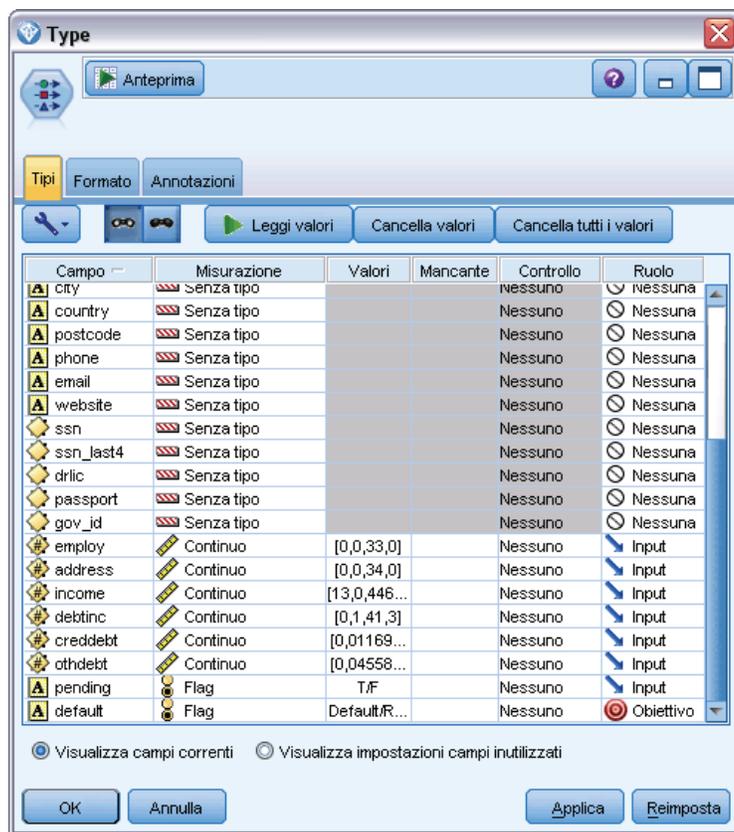
Figura 4-1
Stream iniziale con nodo Modelli



Oltre ai dettagli dei prestiti concessi in passato, l'insieme di dati *loan_applications.csv* include i dettagli di 150 clienti le cui richieste di prestito sono ancora in sospeso, con un totale di 850 record.

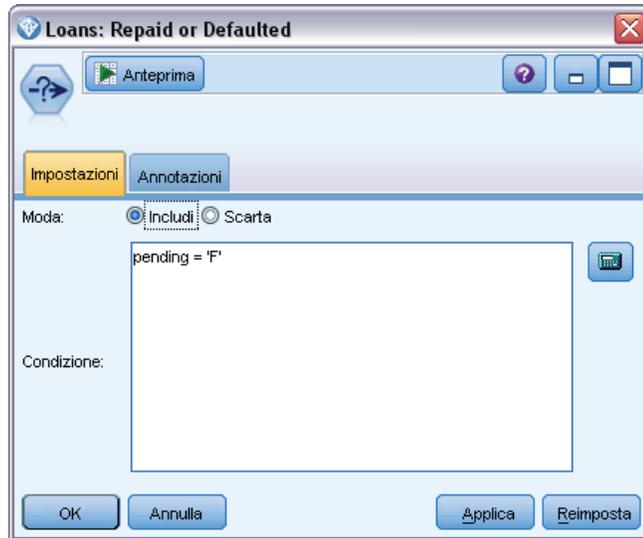
Non tutti i campi dell'insieme di dati sono utili per effettuare la previsione, ad esempio i campi dei nomi possono essere ignorati. Il nodo Tipo esclude i campi da ignorare impostando il loro ruolo su Nessuno. Per i campi da utilizzare per effettuare la previsione, il ruolo è impostato su Input, mentre per i campi di cui il modello tenta di prevedere il valore, il ruolo è impostato su Obiettivo.

Figura 4-2
Ruoli dei campi impostati nel nodo Tipo



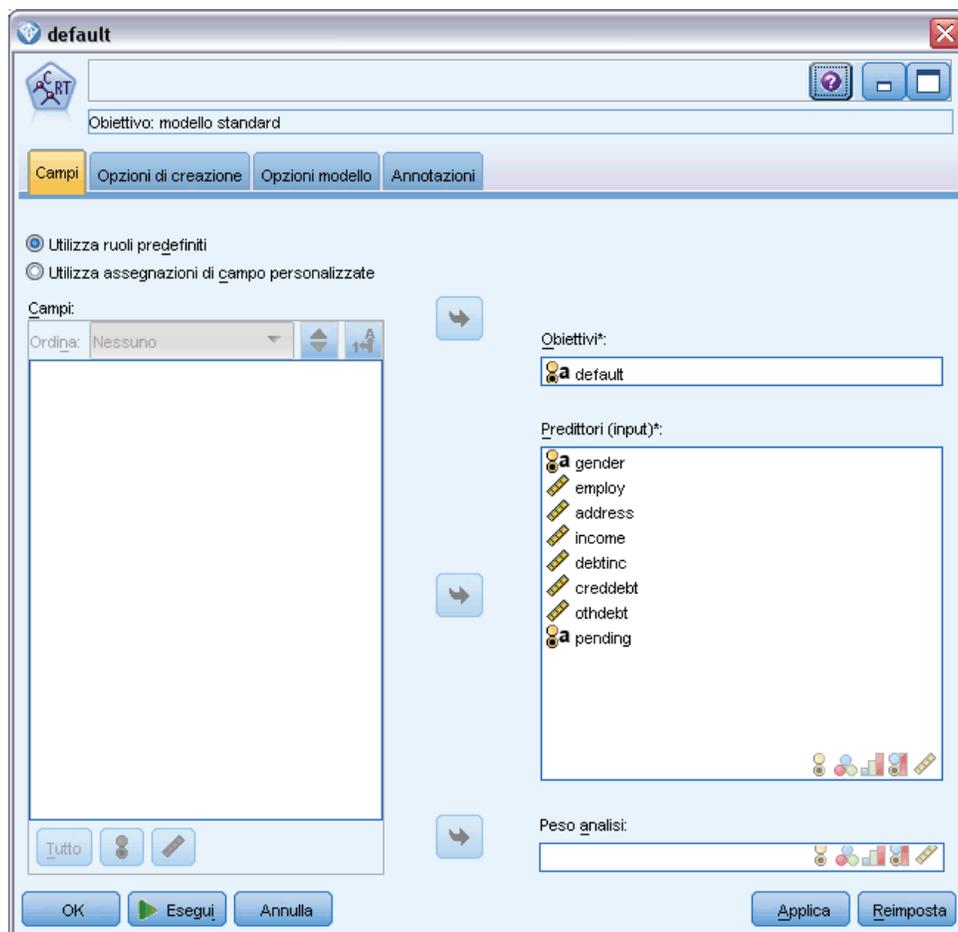
Poiché il modello deve effettuare le previsioni unicamente in base ai dati passati, lo stream include un nodo Seleziona che include soltanto i prestiti che *non* sono contrassegnati come in sospeso, eliminando così i 150 prestiti in sospeso.

Figura 4-3
Eliminazione delle richieste di prestito in sospeso



Con i prestiti in sospeso eliminati, solo i dettagli dei rimanenti 700 prestiti che sono stati rimborsati o non rimborsati per inadempienza passano al nodo Modelli. La banca può avere utilizzato uno dei diversi algoritmi di SPSS Modeler per creare un buon modello. In questo caso, è stato usato un nodo C&R Tree, che verrà utilizzato per creare un modello che preveda quali sono i clienti probabilmente inadempienti in base alle prestazioni passate dei clienti della banca.

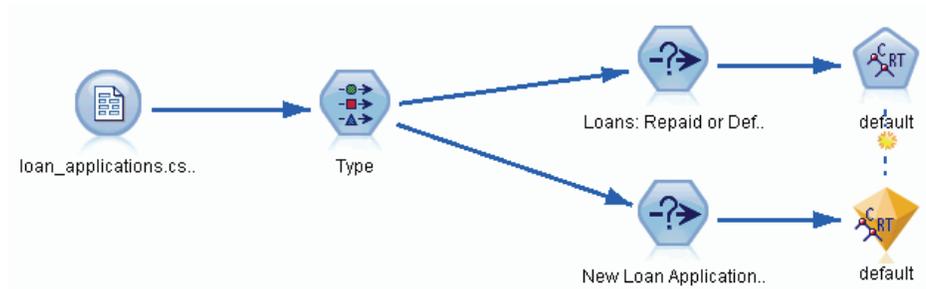
Figura 4-4
Assegnazione di campi predittori e obiettivi



I campi utilizzati per effettuare la previsione sono definiti campi predittori e il campo di cui il modello tenta di prevedere il valore, *default* in questo caso, viene impostato come campo obiettivo, come definito in precedenza dal nodo Tipo.

L'esecuzione di questo stream genera un insieme di modelli contenente il modello creato a partire dai campi predittori.

Figura 4-5
Stream con aggiunta dell'insieme di modelli



Ora l'analista della banca può utilizzare il modello per iniziare a prevedere se i clienti con rimborsi in sospeso probabilmente saranno inadempienti. Utilizzando l'insieme di dati originale, l'analista inserisce un nodo Selezione che questa volta include solo i 150 record di prestiti contrassegnati come in sospeso anziché eliminarli. L'analista passa i record direttamente al modello, aggiungendo un nodo Distribuzione per la rappresentazione visiva delle previsioni del modello.

Figura 4-6

Stream con la selezione delle nuove richieste di prestito e l'aggiunta del nodo Distribuzione



Il nodo Distribuzione mostra la distribuzione dei valori del campo *\$R-default* nel modello. Questo campo viene aggiunto al modello dati dal nodo C&R Tree quando viene eseguito. Il campo contiene la previsione se ogni nuovo richiedente rimborserà il prestito o sarà insolvente e verrà utilizzato in un secondo tempo per confrontare l'effetto dell'aggiunta dell'analisi delle entità.

Eseguendo questa parte dello stream, l'analista sa, grazie all'output del nodo Distribuzione, che la previsione è di 137 su 150 nuovi richiedenti che dovrebbero rimborsare i prestiti. I rimanenti 13 probabilmente saranno inadempienti, quindi l'analista consiglierà alla banca di rifiutare le loro richieste.

Figura 4-7

Output del nodo Distribuzione senza analisi delle entità

Valore	Proporzione	%	Conteggio
Default		8.67	13
Repaid		91.33	137

Aggiunta dell'analisi delle entità

A questo punto si può vedere se la situazione può essere migliorata aggiungendo l'analisi delle entità all'equazione. Immaginiamo che uno specialista di analisi delle entità venga chiamato dalla banca per verificare se sono state inserite voci fraudolente nei record dei clienti nella sorgente dati. È possibile che vi siano record duplicati a causa di errori di inserimento dei dati, ma è anche possibile che un cliente che ha richiesto un prestito stia tentando di camuffare la propria identità. In ogni caso, la banca deve avere il quadro corretto della situazione.

Ai fini di questo esempio, supponiamo che sia già stato creato un repository di entità. Per ulteriori informazioni, vedere l'argomento [Creazione del repository](#) in il capitolo 3 a pag. 16.

Inserimento dei dati di origine nel repository

Per prima cosa, è necessario aggiungere un nodo Esportazione EA al nodo della sorgente dati per esportare i dati di input nel repository delle entità.

Figura 4-8

Aggiunta di un nodo Esportazione EA al nodo della sorgente dati



Per poter esportare i dati, è necessario prima mappare i campi della sorgente dati alle funzioni nel repository di entità. Questa operazione è necessaria poiché diverse sorgenti dati possono utilizzare diversi nomi di campi per lo stesso tipo di informazioni. Il repository di entità fornisce un insieme standard di tipi di informazioni (dette “funzioni”) per evitare duplicati.

Figura 4-9

Mappatura dei campi alle funzioni

The screenshot shows the 'Entity Analytics' window with the 'Input' tab selected. The configuration is as follows:

- Modalità: Aggiungi al repository Svuota repository prima dell'esportazione
- Repository entità: dba@REPOS1
- Tag sorgente: TEST Tipo di entità: PERSON
- Chiave univoca: key

Campo	Mappato a funzione	Utilizzo
key		(Automatico)
fname	NAME.GIVEN_NAME	(Automatico)
mname	NAME.MIDDLE_NAME	(Automatico)
lname	NAME.SUR_NAME	(Automatico)
generation	NAME.NAME_GEN	(Automatico)
dob	DOB.DOB	(Automatico)
dod		(Automatico)
gender	GENDER.GENDER	(Automatico)
addr1	ADDRESS.ADDR1	(Automatico)

Buttons at the bottom: OK, Esegui, Annulla, Importa mappatura..., Esporta mappatura..., Applica, Reimposta.

Nel nodo Esportazione EA, impostare i dettagli relativi al repository: i dettagli della connessione, il tag sorgente (per identificare la sorgente dati, TEST in questo caso), il tipo di entità (l'insieme di funzioni utilizzato, in questo caso PERSON) e il campo della chiave univoca (per identificare in modo univoco ogni record). In questo caso, utilizzare il campo key come chiave univoca.

Ora è possibile impostare le mappature. Nell'insieme di funzioni in uso, alcune funzioni corrispondono ai campi *fname*, *mname*, *lname*, *generation*, *dob*, *gender*, *addr1*, *city*, *country*, *postcode*, *phone*, *email*, *ssn*, *drlic* e *passport*.

Iniziare con l'impostazione della mappatura per *fname*. Fare doppio clic sulla colonna Mappato a funzione della tabella nella riga *fname*, scorrere verso il basso fino alla voce NAME.GIVEN_NAME e fare clic per creare la mappatura.

Ora mappare i campi rimanenti che hanno funzioni corrispondenti, in modo tale che l'intero insieme di mappature sia simile a quello riportato di seguito.

Tabella 4-1
Campi mappati alle funzioni del repository

Campo	Mappato a funzione
<i>fname</i>	NAME.GIVEN_NAME
<i>mname</i>	NAME.MIDDLE_NAME
<i>lname</i>	NAME.SUR_NAME
<i>generation</i>	NAME.NAME_GEN
<i>dob</i>	DOB.DOB
<i>gender</i>	GENDER.GENDER
<i>addr1</i>	ADDRESS.ADDR1
<i>city</i>	ADDRESS.CITY
<i>country</i>	ADDRESS.COUNTRY
<i>postcode</i>	ADDRESS.POSTAL_CODE
<i>phone</i>	PHONE.PHONE_NUM
<i>email</i>	EMAIL_ADDR.ADDR
<i>ssn</i>	SSN.ID_NUM
<i>drlic</i>	DRLIC.ID_NUM
<i>passport</i>	PASSPORT.ID_NUM

Fare clic su Esegui per esportare i dati nel repository. L'operazione richiede un certo tempo, quindi quando la finestra di dialogo Feedback di esecuzione si chiude, l'esportazione è completata.

Letture delle identità risolte

Quando si esportano i dati nel repository, il sistema di analisi delle entità inizia a risolvere i possibili conflitti di identità, assegnando alle entità un identificatore univoco che in seguito apparirà come campo *\$EA-ID*. (Nota: non si tratta del campo Chiave univoca del nodo Esportazione EA, che si utilizza per identificare in modo univoco i record della sorgente dati.

Il primo passaggio per la lettura delle identità risolte è aggiungere un nodo di input Entity Analytics(EA) allo stream. Tale nodo non deve essere connesso ad altro in questa fase.

Figura 4-10
Nodo di input Entity Analytics(EA)



Entity Analytics(EA)

Aprire il nodo di input Entity Analytics(EA) e impostare i dettagli del repository di entità. Viene visualizzato un elenco delle sorgenti dati esportate nel repository, in questo caso è una sola.

Figura 4-11
Selezione di una sorgente dati nel repository



Selezionare la casella di controllo per la sorgente dati TEST e fare clic su OK.

Osserviamo come ha agito sui dati il sistema di analisi delle entità. Collegare un nodo Tabella al nodo di input Entity Analytics(EA), aprire il nodo Tabella e fare clic su Esegui per visualizzare la finestra di output del nodo Tabella.

Figura 4-12
Output del nodo Tabella

	\$EA-ID	\$EA-SRC	\$EA-KEY
1	1	TEST	REF00000001
2	2	TEST	REF00000002
3	3	TEST	REF00000003
4	4	TEST	REF00000004
5	5	TEST	REF00000005
6	6	TEST	REF00000006
7	7	TEST	REF00000007
8	8	TEST	REF00000008
9	9	TEST	REF00000009
10	10	TEST	REF00000010
11	11	TEST	REF00000011
12	12	TEST	REF00000012
13	13	TEST	REF00000013
14	14	TEST	REF00000014
15	15	TEST	REF00000015
16	16	TEST	REF00000016
17	17	TEST	REF00000017
18	18	TEST	REF00000018
19	19	TEST	REF00000019
20	20	TEST	REF00000020

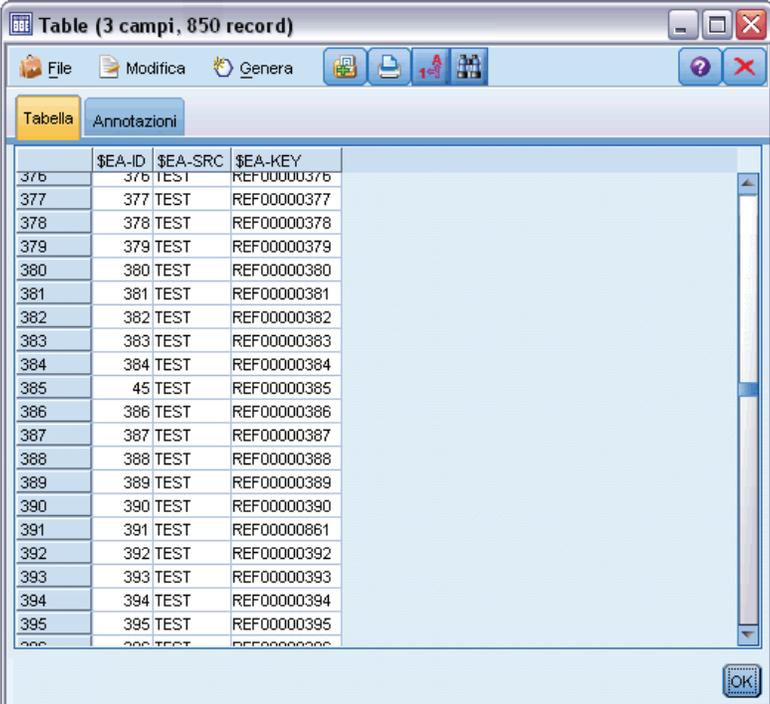
Solo un campo risulta familiare, quello con etichetta *\$EA-KEY*. Si tratta in realtà del campo *key* della sorgente dati e risulta poiché è stato scelto come campo Chiave univoca nel nodo Esportazione EA.

Il sistema ha aggiunto altri due campi. Il campo *\$EA-ID* è l'identificatore univoco, non dei record sorgente ma delle identità risolte. Tra poco vedremo la differenza. Il campo *\$EA-SRC* identifica la sorgente dei dati; qui indica TEST poiché è il tag sorgente assegnato al campo nel nodo Esportazione EA.

Che cosa è accaduto a tutti gli altri campi dei dati di origine? Nulla, sono ancora nel repository, è solo per migliorare le prestazioni che il nodo di input Entity Analytics(EA) passa solo l'insieme minimo di campi a valle per un'ulteriore elaborazione.

A questo punto, far scorrere verso il basso l'output del nodo Tabella fino alla riga 385.

Figura 4-13
Differenze tra le righe di output di Tabella e i numeri di \$EA-ID

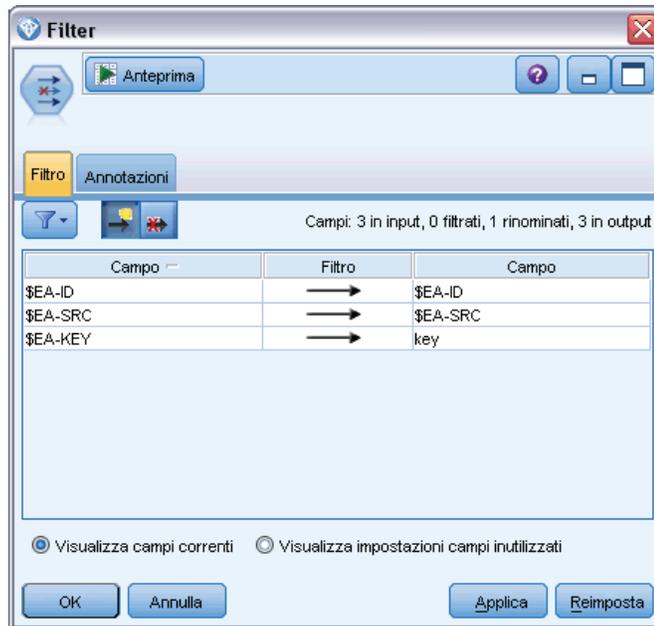


	\$EA-ID	\$EA-SRC	\$EA-KEY
375	375	TEST	REF00000375
377	377	TEST	REF00000377
378	378	TEST	REF00000378
379	379	TEST	REF00000379
380	380	TEST	REF00000380
381	381	TEST	REF00000381
382	382	TEST	REF00000382
383	383	TEST	REF00000383
384	384	TEST	REF00000384
385	45	TEST	REF00000385
386	386	TEST	REF00000386
387	387	TEST	REF00000387
388	388	TEST	REF00000388
389	389	TEST	REF00000389
390	390	TEST	REF00000390
391	391	TEST	REF00000861
392	392	TEST	REF00000392
393	393	TEST	REF00000393
394	394	TEST	REF00000394
395	395	TEST	REF00000395

Notare come il numero *\$EA-ID* sembra essere fuori sequenza. Il sistema di analisi delle entità ha stabilito che il record REF00000385 fa riferimento alla persona identificata come entità 45, che detiene anche il record REF00000045. Scorrendo ancora verso il basso l'output, si trovano altri numeri fuori sequenza, ad esempio alle righe 485, 517, 520 e così via. È necessario approfondire.

Per prima cosa, si rileva il fatto che il campo *\$EA-KEY* contiene i dati del campo *key* nella sorgente dati; è quindi necessario rinominarlo in *key*. Collegare un nodo Filtro al nodo di input Entity Analytics(EA) e aprire il nodo Filtro. Fare doppio clic sulla stringa *\$EA-KEY* nella seconda colonna Campo e digitare *key*.

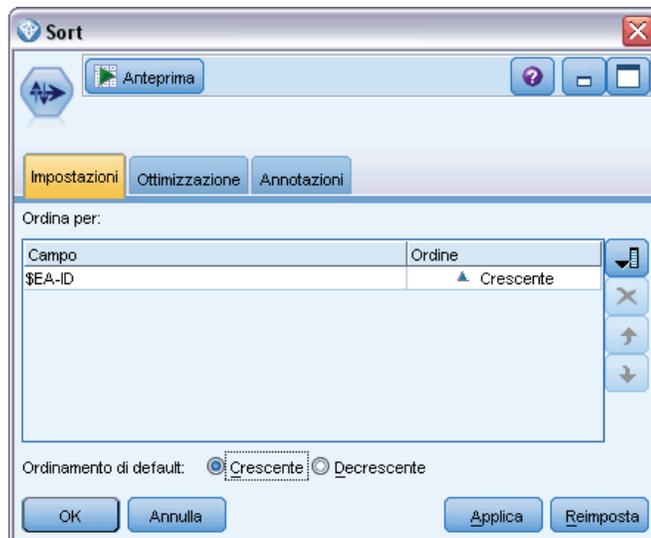
Figura 4-14
Ridenominazione del campo \$EA-KEY



Fare clic su OK per chiudere il nodo Filtro.

Ora è necessario ordinare gli ID entità *\$EA-ID* in ordine crescente. Collegare un nodo Ordina al nodo Filtro. Aprire il nodo Ordina, fare clic sul primo pulsante in alto accanto alla tabella Ordina per, selezionare *\$EA-ID* e fare clic su OK.

Figura 4-15
Ordinamento degli ID entità in ordine crescente



Lasciare selezionato l'ordine crescente e fare clic su OK.

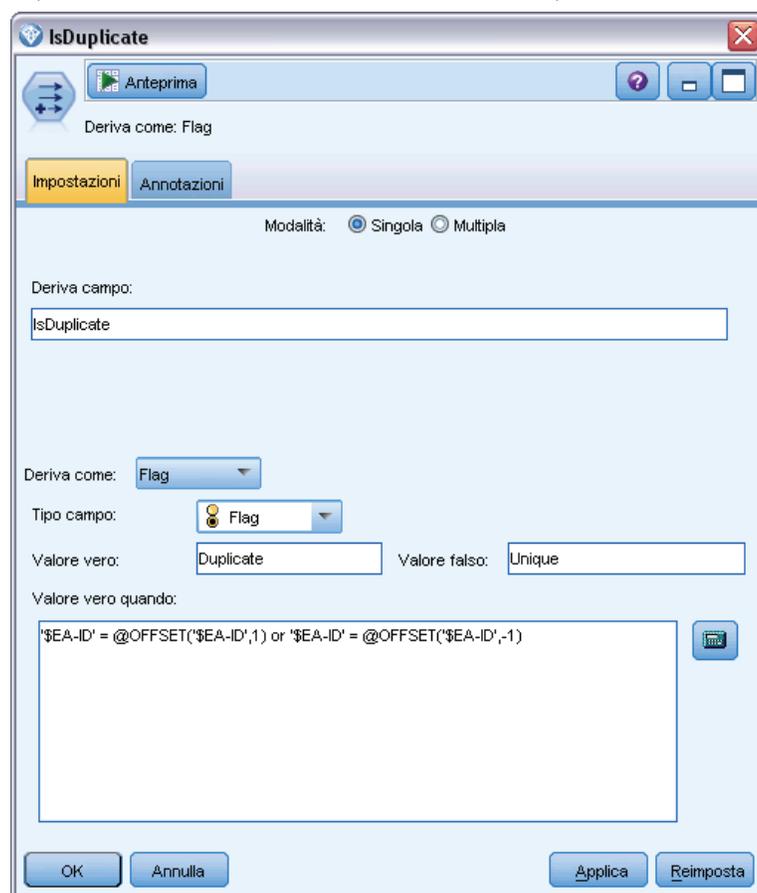
Ora è necessario creare un campo extra che indichi se un record è univoco o duplicato. Collegare un nodo Nuovo campo al nodo Ordina. Aprire il nodo Nuovo campo e impostare il nome di Deriva campo su *IsDuplicate*. Dall'elenco Deriva come, scegliere *Flag*, che imposta anche il Tipo campo su *Flag*. Impostare il campo Valore vero su *Duplicate* e il campo Valore falso su *Unique*.

Per trovare i record duplicati, si utilizza una funzione di sequenza speciale, *@OFFSET*, fornita con SPSS Modeler.

Immettere quanto segue nel campo If:

```
'$EA-ID' = @OFFSET('$EA-ID',1) oppure '$EA-ID' = @OFFSET('$EA-ID',-1)
```

Figura 4-16
Impostazione della condizione nel nodo Nuovo campo



Con gli ID entità in ordine crescente, la funzione *@OFFSET* verifica se due ID entità adiacenti sono identici e, se lo sono, significa che i record sono duplicati. In questo caso, il relativo valore *IsDuplicate* è impostato su *Duplicate*, in caso contrario su *Unique*.

Fare clic su OK per chiudere il nodo.

Per vedere gli effetti del nodo Nuovo campo, allegare un nodo Tabella al nodo Nuovo campo, aprire il nodo Tabella e fare clic su Esegui. Far scorrere verso il basso la finestra dell'output del nodo Tabella fino alla riga 45.

Figura 4-17
Output del nodo Nuovo campo

	\$EA-ID	\$EA-SRC	key	IsDuplicate
39	39	TEST	REF00000039	Unique
40	40	TEST	REF00000040	Unique
41	41	TEST	REF00000041	Unique
42	42	TEST	REF00000042	Unique
43	43	TEST	REF00000043	Unique
44	44	TEST	REF00000044	Unique
45	45	TEST	REF00000045	Duplicate
46	45	TEST	REF00000385	Duplicate
47	46	TEST	REF00000046	Unique
48	47	TEST	REF00000047	Unique
49	48	TEST	REF00000048	Unique
50	49	TEST	REF00000049	Unique
51	50	TEST	REF00000050	Unique
52	51	TEST	REF00000051	Unique
53	52	TEST	REF00000052	Unique
54	53	TEST	REF00000053	Unique
55	54	TEST	REF00000054	Unique
56	55	TEST	REF00000055	Unique
57	56	TEST	REF00000056	Unique
58	57	TEST	REF00000057	Unique

Si può fare un confronto con l'output generato direttamente dal nodo di input Entity Analytics(EA). Il sistema aveva già rilevato che il record REF00000385 fa riferimento alla stessa persona dell'entità 45. Ora siamo in una fase successiva della procedura e abbiamo verificato che i record REF00000045 e REF00000385 sono duplicati poiché fanno entrambi riferimento all'entità 45.

Far scorrere ulteriormente verso il basso la finestra dell'output per vedere gli altri record contrassegnati come duplicati.

Per ottenere un report contenente i record duplicati, collegare un nodo Report (dalla scheda Output della palette dei nodi) al nodo Nuovo campo *IsDuplicate*. Aprire il nodo Report, copiare il testo seguente nel campo di input della scheda Modello e fare clic su Esegui.

```
<html>
<h1>List of duplicate customer records.

<h2>This report was generated: [@TODAY]

<h2>Duplicate records
<table>
<tr>
<td>Entity ID</td>
```

```
<td>Key</td>
</tr>

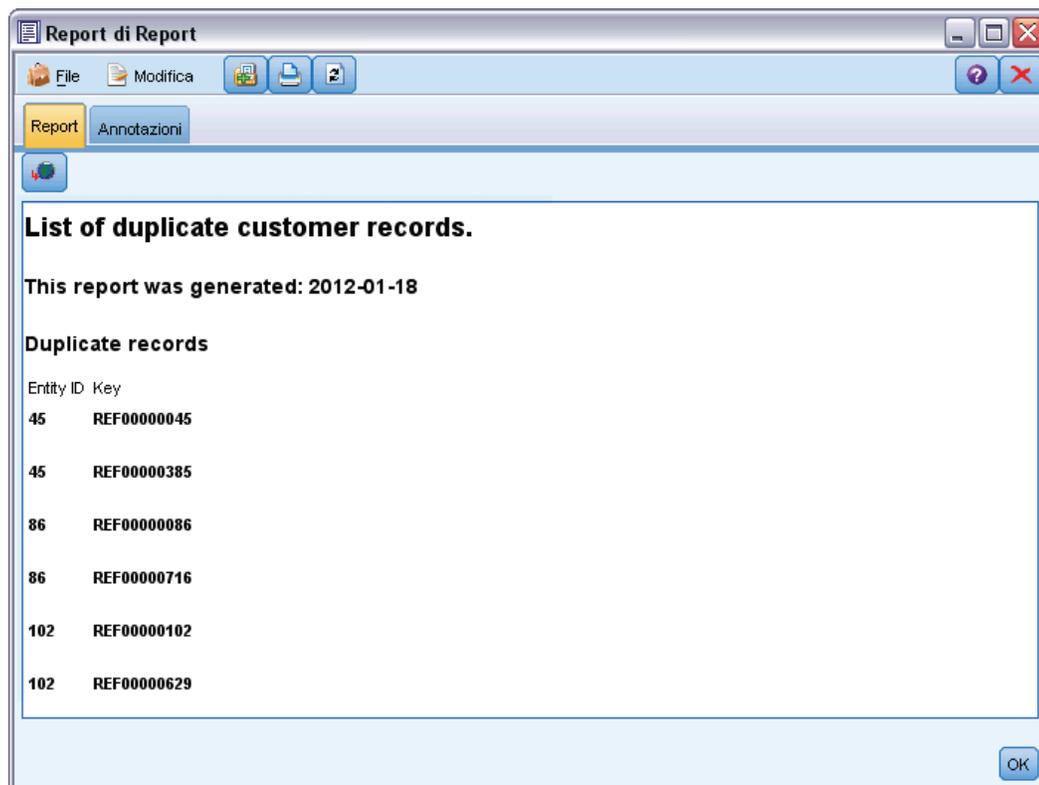
#WHERE IsDuplicate = "Duplicate"

<tr>
  <td>['$EA-ID']</td>
  <td>[key]</td>
</tr>
#
</table>

</html>
```

Viene generato il seguente output.

Figura 4-18
Output del nodo report



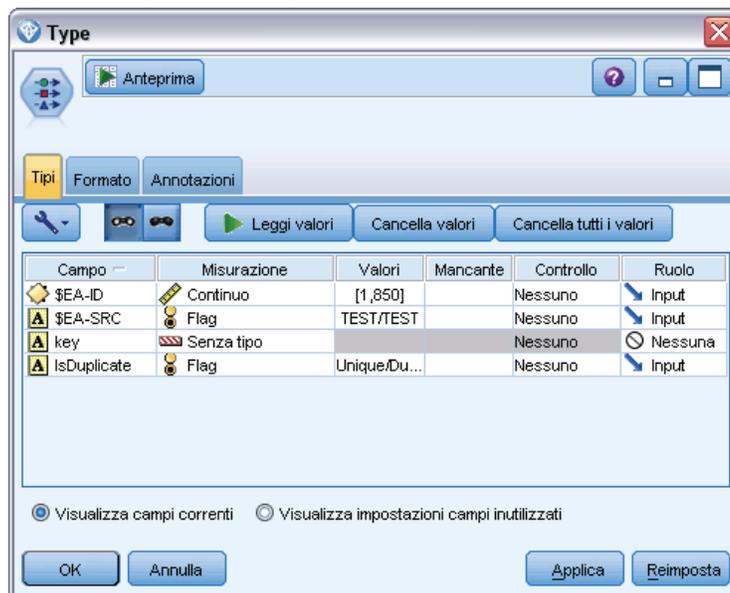
Il report in questo caso utilizza il formato HTML anche se è possibile utilizzare anche la formattazione XML o ASCII.

Confronto dell'output dell'analisi delle entità con il modello originale

La fase finale dell'esempio è vedere se aggiungendo l'analisi delle entità si rilevano differenze rispetto alla previsione originale della banca. Come sappiamo, il modello originale aveva previsto 13 clienti insolventi su 150 richieste di prestito in sospeso. Verrà utilizzato un nodo Unione per unire l'output di quel modello con le informazioni sui record duplicati di Entity Analytics per vedere se in questo modo la previsione cambia.

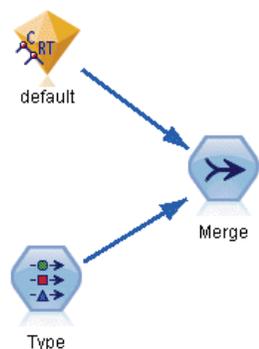
Per prima cosa, è necessario assicurarsi che i nuovi campi aggiunti dall'analisi delle entità possiedano i tipi di dati corretti, o *livelli di misurazione*, come sono chiamati in SPSS Modeler. Collegare un nodo Tipo al nodo Nuovo campo IsDuplicate, aprire il nodo Tipo e fare clic sul pulsante Leggi valori.

Figura 4-19
Impostazioni per il nodo Tipo



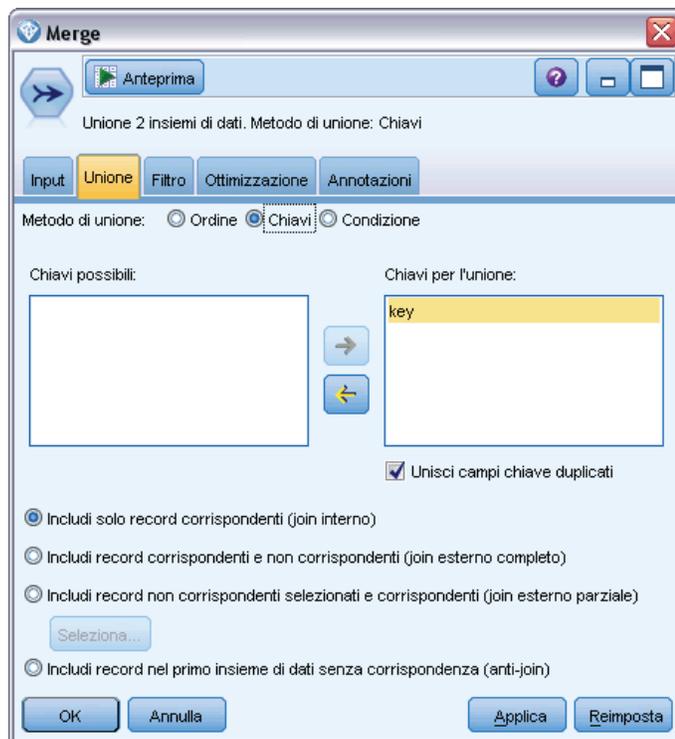
Ora è possibile aggiungere il nodo Unione. Collegarlo al nodo Tipo e collegarlo anche all'insieme di modelli che contiene il modello originale. A questo scopo, fare clic con il pulsante destro del mouse sull'insieme di modelli, scegliere Connetti, quindi fare clic sul nodo Unione, che ora dovrebbe apparire con due frecce di input.

Figura 4-20
Input per il nodo Unione



Aprire il nodo Unione, impostare il Metodo di unione su Chiavi e fare clic sul pulsante con la freccia a destra per spostare il campo key da Chiavi possibili a Chiavi per l'unione, quindi fare clic su OK.

Figura 4-21
Indicazione del campo Chiavi per l'operazione di unione



Il confronto è quasi terminato. Tuttavia, se si stava collegando un nodo Distribuzione e lo si esegue in questo punto, non si visualizza alcuna modifica rispetto alla previsione originale. Benché lo stream unisca ora l'output dell'insieme di modelli originale con i nuovi campi creati dall'analisi delle entità, il campo di previsione stesso (*\$R-default*) nel modello dati non è stato aggiornato con le nuove informazioni.

Pertanto, dovrà essere utilizzato un nodo Riempimento per sostituire i valori del campo. Collegare un nodo Riempimento al nodo Unione e aprire il nodo Riempimento.

Fare clic sul primo pulsante in alto a destra di Riempi campi, scorrere fino in fondo all'elenco, scegliere \$R-default e fare clic su OK. Questo è il campo i cui valori devono essere modificati se la condizione specificata nel resto della finestra di dialogo è soddisfatta.

Per specificare la condizione, verificare che l'opzione Sostituisci sia impostata su In base alla condizione, quindi inserire quanto segue nel campo Condizione:

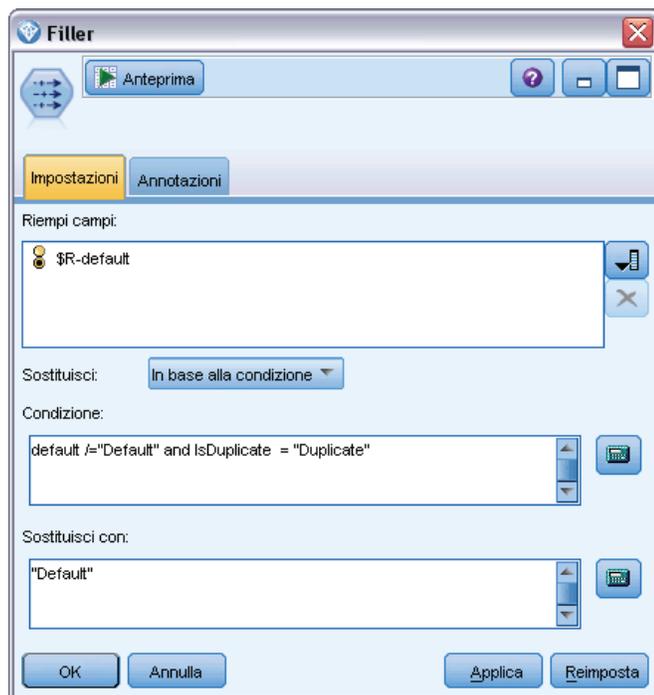
```
default /="default" and IsDuplicate = "Duplicate"
```

Nel campo Sostituisci con, inserire:

```
"default"
```

Figura 4-22

Indicazione della condizione per la sostituzione dei valori dei campi



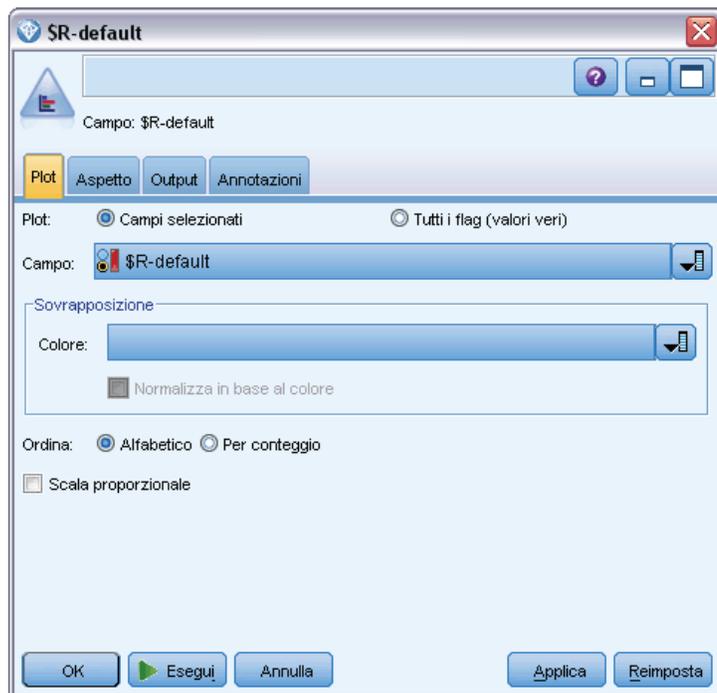
Per queste impostazioni è necessaria una breve spiegazione. La condizione stabilisce che per ogni record in cui il valore del campo di *default* dell'insieme di dati originale non è uguale a quello di Default e il record è contrassegnato come duplicato, il valore del campo *\$R-default* del modello debba essere impostato su Default.

Il campo *\$R-default* è il campo del modello che contiene la previsione circa il fatto che un cliente possa essere insolvente per il prestito. In questo modo, i clienti con record duplicati vengono aggiunti al modello come potenziali insolventi.

Fare clic su OK per chiudere il nodo Riempimento.

A questo punto è possibile vedere la differenza utilizzando Entity Analytics. Dalla palette Grafici, collegare un nodo Distribuzione al nodo Riempimento e aprire il nodo Distribuzione. Fare clic sull'elenco Campo e scegliere \$R-default.

Figura 4-23
Impostazioni per il nodo Distribuzione



Fare clic su Esegui per creare il grafico della nuova previsione.

Figura 4-24
Output del nodo Distribuzione dopo l'analisi delle entità

Valore	Proporzione	%	Conteggio
Default	10.67	10.67	16
Repaid	89.33	89.33	134

Ora sono presenti 16 richieste a rischio anziché 13. Queste richieste aggiuntive possono incidere molto in caso di insolvenza, quindi è possibile dimostrare graficamente alla banca il vantaggio offerto dall'utilizzo di Entity Analytics per le operazioni di valutazione del rischio.

Riepilogo

Questo esempio ha indicato come, utilizzando l'analisi delle entità, sia possibile eliminare la duplicazione dei record nei dati relativi a persone o organizzazioni, migliorando la qualità delle previsioni.

Nota: idealmente i record duplicati vanno eliminati prima di qualsiasi altra elaborazione. La procedura può essere seguita dall'utilizzo di un nodo Preparazione automatica dati (ADP) per analizzare i dati e individuare le correzioni, escludere i campi problematici o probabilmente inutili e derivare all'occorrenza nuovi attributi, migliorando le performance grazie allo screening intelligente.

La combinazione dell'analisi delle entità e della preparazione automatica dei dati assicura la massima pulizia dei dati utilizzati.

Note

Queste informazioni sono state preparate per prodotti e servizi offerti in tutto il mondo.

IBM potrebbe non offrire i prodotti, i servizi o le funzionalità di cui si tratta nel presente documento in altri paesi. Contattare il rappresentante IBM locale per informazioni sui prodotti e i servizi attualmente disponibili nella propria zona. Qualsiasi riferimento a un prodotto, programma o servizio IBM non intende dichiarare o implicare che sia possibile utilizzare esclusivamente tale prodotto, programma o servizio IBM. Potrà invece essere utilizzato qualsiasi prodotto, programma o servizio con funzionalità equivalente e che non violi i diritti di proprietà intellettuale di IBM. Tuttavia, è responsabilità dell'utente valutare e verificare il funzionamento di qualsiasi prodotto, programma o servizio non IBM.

IBM può essere titolare di brevetti o domande di brevetto relativi alla materia oggetto del presente documento. La consegna del presente documento non conferisce alcuna licenza rispetto a questi brevetti. Rivolgere per iscritto i quesiti sulle licenze a:

IBM Director of Licensing, IBM Corporation, North Castle Drive, Armonk, NY 10504-1785, U.S.A.

Per richieste di informazioni sulle licenze riguardanti il set di caratteri a byte doppio (DBCS), contattare l'Intellectual Property Department di IBM del proprio paese, oppure inviare le richieste in forma scritta all'indirizzo:

Intellectual Property Licensing, Legal and Intellectual Property Law, IBM Japan Ltd., 1623-14, Shimotsuruma, Yamato-shi, Kanagawa 242-8502 Giappone.

Il seguente paragrafo non si applica per il Regno Unito o altri paesi in cui le presenti disposizioni non sono conformi alle leggi locali: INTERNATIONAL BUSINESS MACHINES FORNISCE QUESTA PUBBLICAZIONE “COSÌ COM'È” SENZA GARANZIA DI ALCUN TIPO, SIA ESSA ESPRESSA O IMPLICITA, INCLUSE, MA NON LIMITATE A, LE GARANZIE IMPLICITE DI NON VIOLAZIONE, COMMERCIALIZZABILITÀ O IDONEITÀ A UNO SCOPO SPECIFICO. Alcuni stati non consentono limitazioni di garanzie espresse o implicite in determinate transazioni, pertanto quanto sopra potrebbe non essere applicabile.

Le presenti informazioni possono includere imprecisioni tecniche o errori tipografici. Le modifiche periodiche apportate alle informazioni contenute in questa pubblicazione verranno inserite nelle nuove edizioni della pubblicazione. IBM può apportare miglioramenti e/o modifiche al/ai prodotto/i e/o al/ai programma/i descritti nella presente pubblicazione in qualsiasi momento senza preavviso.

Qualsiasi riferimento nelle presenti informazioni a siti Web non IBM viene fornito esclusivamente per facilitare la consultazione e non rappresenta in alcun modo un'approvazione o sostegno da parte nostra di tali siti Web. I materiali contenuti in tali siti Web non fanno parte dei materiali di questo prodotto IBM e il loro utilizzo è esclusivamente a rischio dell'utente.

IBM può utilizzare o distribuire eventuali informazioni fornite dall'utente nei modi che ritiene appropriati senza incorrere in alcun obbligo nei confronti dell'utente.

I licenziatari del programma che desiderassero informazioni su di esso allo scopo di abilitare: (i) lo scambio di informazioni tra programmi creati indipendentemente e altri programmi (questo compreso) e (ii) l'utilizzo in comune delle informazioni scambiate, dovranno rivolgersi a:

IBM Software Group, All'attenzione di: Licensing, 233 S. Wacker Dr., Chicago, IL 60606, USA.

Tali informazioni saranno fornite in conformità ai termini e alle condizioni in vigore e, in alcuni casi, dietro pagamento.

Il programma concesso in licenza descritto nel presente documento e tutto il materiale correlato disponibile sono forniti da IBM in base ai termini del contratto di licenza cliente IBM, del contratto di licenza internazionale IBM o del contratto equivalente esistente tra le parti.

Tutti i dati sulle prestazioni qui contenuti sono stati elaborati in ambiente controllato. Di conseguenza, i risultati ottenuti con sistemi operativi diversi possono variare in modo significativo. Alcune misurazioni potrebbero essere state effettuate su sistemi in corso di sviluppo e non c'è garanzia che tali misurazioni coincidano con quelle effettuate sui sistemi comunemente disponibili. Inoltre, alcune misurazioni potrebbero essere stime elaborate tramite l'estrapolazione. I risultati effettivi potrebbero variare. Gli utenti di questo documento devono verificare i dati relativi al proprio ambiente specifico.

Le informazioni relative a prodotti non IBM sono state ottenute dai fornitori di tali prodotti, da loro annunci pubblicati e da altre fonti disponibili al pubblico. IBM non ha verificato tali prodotti e non può confermare l'accuratezza delle prestazioni, la compatibilità o qualsiasi altra dichiarazione relativa a prodotti non IBM. Eventuali domande in merito alle funzionalità dei prodotti non IBM vanno indirizzate ai fornitori di tali prodotti.

Qualsiasi affermazione relativa agli obiettivi e alla direzione futura di IBM è soggetta a modifica o revoca senza preavviso e concerne esclusivamente gli scopi dell'azienda.

Le presenti informazioni includono esempi di dati e report utilizzati in operazioni aziendali quotidiane. Per fornire una descrizione il più possibile esaustiva, gli esempi includono nomi di persone, società, marchi e prodotti. Tutti questi nomi sono fittizi e ogni somiglianza a nomi e indirizzi utilizzati da aziende reali è puramente casuale.

Per chi visualizza queste informazioni a video: le fotografie e le illustrazioni a colori potrebbero non essere disponibili.

Marchi

IBM, il logo IBM, ibm.com e SPSS sono marchi di IBM Corporation, registrati in numerose giurisdizioni nel mondo. Un elenco aggiornato dei marchi IBM è disponibile sul Web all'indirizzo <http://www.ibm.com/legal/copytrade.shtml>.

Intel, il logo Intel, Intel Inside, il logo Intel Inside, Intel Centrino, il logo Intel Centrino, Celeron, Intel Xeon, Intel SpeedStep, Itanium e Pentium sono marchi o marchi registrati di Intel Corporation o delle sue consociate negli Stati Uniti e in altri paesi.

Linux è un marchio registrato di Linus Torvalds negli Stati Uniti e/o negli altri paesi.

Microsoft, Windows, Windows NT e il logo Windows sono marchi di Microsoft Corporation negli Stati Uniti e/o negli altri paesi.

UNIX è un marchio registrato di The Open Group negli Stati Uniti e in altri paesi.

Java e tutti i marchi e i logo basati su Java sono marchi di Sun Microsystems, Inc. negli Stati Uniti e/o negli altri paesi.

Altri nomi di prodotti e servizi possono essere marchi commerciali di IBM o di altre aziende.



- analisi entità
 - confronto con analisi predittiva, 2
 - definizione, 1
 - utilizzo con IBM SPSS Modeler, 4
- assegnazioni delle porte
 - configurazione per l'analisi delle entità, 40
- caratteristiche
 - repository entità, 7, 19–21, 23–24, 26, 37
- chiavi univoche
 - analisi entità, 7
 - repository entità, 20
- configurazione
 - repository entità, 22, 31–32
- corrispondenza entità, impostazione soglia, 30
- eliminazione
 - repository entità, 43–44
- esportazione
 - dati a un repository di entità, 8
- identità risolte, analisi, 32
- identità, risoluzione con Entity Analytics, 8
- informazioni sui tipi, impostazione per l'analisi delle entità, 34
- mappatura di campi
 - a funzioni del repository entità, 7, 19–21, 23, 37
- marchi, 66
- nodi
 - aggiunta a uno stream di analisi delle entità, 34
- nodi di esportazione
 - Entity Analytics, 8
- nodi di input
 - Entity Analytics, 11
- note legali, 65
- repository
 - amministrazione dell'analisi delle entità, 40
 - analisi entità, 6–8, 14, 16–24, 26, 28, 31–32, 37, 43–44
- repository entità, 14
 - attività amministrative, 40
 - caratteristiche, 26
 - configurazione, 22, 31–32
 - configurazione delle assegnazioni delle porte, 40
 - connessione a IBM SPSS Modeler, 7
 - creazione, 6, 16–17
 - eliminazione, 43–44
 - gestione, 24
 - impostazione, 14
 - opzioni, 18
 - svuotamento, 43
- ridenominazione
 - campi dati per analisi entità, 33
- risoluzione di identità, Entity Analytics, 8
- sorgente dati, selezione per analisi entità, 32
- sorgenti dati
 - connessione con analisi entità, 5, 15
 - visualizzazione per analisi entità, 18, 38
- svuotamento
 - repository entità, 43
- tag sorgente
 - repository entità, 20
- tipi di entità
 - analisi entità, 28
 - repository entità, 20
- tipi di utilizzo, analisi entità, 28