

IBM SPSS Modeler Entity Analytics 17
Guida dell'utente

IBM

Nota

Prima di utilizzare queste informazioni e il relativo prodotto, leggere le informazioni in "Note" a pagina 59.

Informazioni sul prodotto

Questa edizione si applica alla versione 17, release 0, modifica 0 di IBM(r) SPSS(r) Modeler ed a tutte le modifiche e release successive se non diversamente indicato nelle nuove edizioni.

Indice

Prefazione	v	Confronto di nuovi casi con il repository (nodo Streaming di EA)	28
Capitolo 1. Analisi delle entità	1	Mappatura dei campi di input alle funzioni (nodo Streaming di EA)	29
L'analisi delle entità	1	Visualizzazione delle mappature dei campi e delle origini dati (nodo Streaming di EA)	30
Analisi delle entità e analisi predittiva	2	Output del nodo Streaming di EA	31
Capitolo 2. Analisi delle entità con IBM SPSS Modeler	5	Utilizzo di IBM SPSS Modeler Entity Analytics con altri prodotti IBM SPSS	31
Utilizzo di analisi delle entità con IBM SPSS Modeler	5	Attività amministrative	32
Fase 1: lettura dei dati di origine in SPSS Modeler	6	Configurazione delle assegnazioni delle porte	32
Fase 2: creazione del repository	6	Gestione delle credenziali dell'amministratore per il database del repository	33
Fase 3: Connessione di SPSS Modeler al repository	7	Spostamento del repository ad una directory di archiviazione diversa	33
Fase 4: Mappatura dei campi di input alle funzioni del repository	7	Impostazioni delle proprietà flusso per i campi data/ora e timestamp	34
Fase 5: esportazione dei dati nel repository e risoluzione delle identità	7	Modifica delle impostazioni di timeout	34
Fase 6: Analisi delle identità risolte	9	Esecuzione di IBM SPSS Modeler Entity Analytics con il client SPSS Modeler e SPSS Modeler Server sul medesimo sistema Windows	35
Fase 7: Risoluzione di nuovi casi rispetto al repository	10	Svuotamento di un repository di entità	35
Fase 8: generazione avvisi	10	Eliminazione di origini dati non utilizzati da un repository	35
Eliminazione di un repository di entità		Eliminazione di un repository quando non è possibile connettersi a esso	36
Capitolo 3. Attività di analisi delle entità	13	Capitolo 4. Funzionamento di Entity Analytics	39
Introduzione	13	Informazioni su questo esempio	39
Impostazione di un repository di entità (nodo Esportazione EA)	13	Il modello originale	39
Il repository di entità	13	Aggiunta dell'analisi delle entità	42
Connessione a un'origine dati	14	Inserimento dei dati di origine nel repository	42
Creazione del repository	14	Lettura delle identità risolte	43
Mappatura dei campi di input alle funzioni (nodo Esportazione EA)	16	Confronto dell'output dell'analisi delle entità con il modello originale	49
Visualizzazione delle mappature dei campi (nodo Esportazione EA)	18	Riepilogo	53
Configurazione di un repository di entità	18	Appendice. Proprietà script per IBM SPSS Modeler Entity Analytics	55
Visualizzazione delle mappature delle sorgenti dati	19	Script con IBM SPSS Modeler Entity Analytics	55
Gestione delle funzioni del repository	20	Proprietà comuni	55
Aggiunta o modifica di una funzione	21	Proprietà entityanalytics_exportnode	55
Anonimizzazione delle funzioni del repository	22	Proprietà entityanalytics_sourcenode	56
Gestione dei tipi di entità	23	Proprietà entityanalytics_processnode	56
Impostazione della soglia per la corrispondenza delle entità	25	Note	59
Riutilizzo di una configurazione di repository	25	Marchi	60
Salvataggio delle modifiche alla configurazione	26	Indice analitico	63
Chiusura della finestra di configurazione	26		
Analisi delle identità risolte (nodo origine Entity Analytics(EA))	26		
Selezione di una sorgente dati	26		
Ridenominazione dei campi dati	27		
Impostazione delle informazioni sui tipi per i campi dati	27		
Aggiunta di nodi al flusso	28		

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 SPSS Modeler invita gli utenti ad applicare la loro competenza specifica nel business il che conduce ad ottenere modelli predittivi più efficaci e a ridurre i tempi di risoluzione dei problemi. SPSS Modeler offre molte tecniche di modellazione come la previsione, la classificazione, la segmentazione e algoritmi di individuazione di 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 di business. 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 di business. Le aziende, le istituzioni e il mondo accademico fanno affidamento sulla tecnologia IBM SPSS per ottenere vantaggi concorrenziali nell'attrarre, mantenere e accrescere i clienti riducendo al contempo i rischi e il pericolo di frodi. Integrando il software IBM SPSS nei processi di tutti i giorni, le aziende diventano aziende predittive in grado di adottare e automatizzare le decisioni per centrare gli obiettivi di business e ottenere vantaggi concorrenziali misurabili. 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 accordo di manutenzione.

Capitolo 1. Analisi delle entità

L'analisi delle entità

IBM SPSS Modeler Entity Analytics aggiunge una dimensione supplementare all'analisi predittiva 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.

Origine 1

Record no.: 70001
Name: Jon Smith
Address: 123 Main Street
Tax Reference: 555-00-1111
Driv. License: 0001133107
Cred. 10229127

Origine

Record no.: 9103
Name: JOHNATHAN Smith
Date of Birth: 06/17/1934
Telephone: 555-1212
Cred. Card: 10229128
Email: jls@mail.com
IP address: 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.

Origine 3

Record no.: 6251
Name: Jon Smith
Telephone: 555-1212
Driv. License: 0001133107
Cred. Card: 10229132

Il numero di patente collega i record nell'origine 1 e 3 mentre il numero di telefono collega le origini 2 e 3. Quindi possiamo essere ragionevolmente sicuri che tutte e tre origini fanno riferimento 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.

Origine 4

Record no.: S45286
Name: John T Smith Jr
Address: 456 Main Street

Telephone: 703-555-2000
Date of birth: 03/12/1984

Record no.: S45287
Name: John T Smith
Address: 456 Main Street
Telephone: 703-555-2000
Driv. License: 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.

Origine 5

Record no.: 769582-2
Name: John T Smith Sr
Address: 456 Main Street
Telephone: 703-555-2000
Driv. License: 009900991
Date of birth: 06/25/1959

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

Caratteristica	Analisi predittiva	Analisi delle entità
Tipi di dati di addestramento	Basato su insiemi e intervalli numerici relativamente piccoli	È possibile sfruttare ampi insiemi (campi senza tipo) come nomi e indirizzi
Dimensione dei dati di addestramento	In genere ignora ampi insiemi (campi senza tipo)	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

Tabella 1. Differenze tra l'analisi predittiva e l'analisi delle entità (Continua).

Caratteristica	Analisi predittiva	Analisi delle entità
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

Capitolo 2. Analisi delle entità con IBM SPSS Modeler

Utilizzo di analisi delle entità con IBM SPSS Modeler

Si sospetta che si possano avere problemi di identità con i dati in uso. Ad esempio, gli individui potrebbero essere più di uno o individui distinti potrebbero sembrare uniti o mancanti. 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 IBM 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 IBM SPSS Modeler. IBM SPSS Modeler è uno strumento di facile uso, basato sulla rappresentazione grafica di un flusso di dati che passa attraverso un numero di nodi. Ogni nodo rappresenta una fase particolare del flusso di lavoro.

IBM SPSS Modeler stesso fornisce una vasta gamma di nodi, che coprono tutte le funzioni standard di data mining. IBM SPSS Modeler Entity Analytics aggiunge nodi per l'utilizzo specifico nelle analisi di entità. Si tratta dei nodi di esportazione EA, del nodo di origine Entity Analytics(EA) e del nodo di elaborazione Streaming di EA.

Il processo è illustrato nella figura seguente.

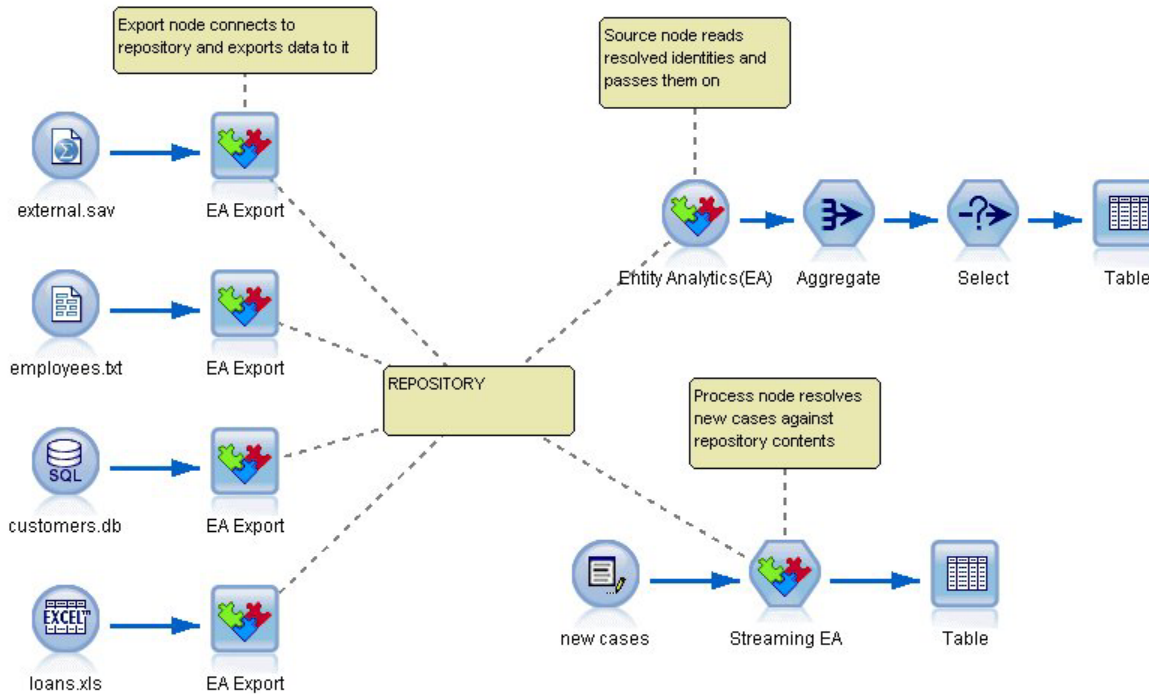


Figura 1. Il processo di analisi delle entità

Fase 1: lettura dei dati di origine in SPSS Modeler

La prima attività consiste nel leggere i dati in SPSS Modeler utilizzando uno o più nodi origine indicati in SPSS Modeler da un'icona rotonda.

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 origine di SPSS Modeler corrispondente. Nell'illustrazione è un nodo origine del database.

Ogni file della sorgente dati deve avere un campo che identifichi in modo univoco ogni record. Se un'origine dati non contiene tale campo, è possibile aggiungerne facilmente uno in SPSS Modeler. Consultare l'argomento "Aggiunta di un identificatore univoco del record" a pagina 14 per ulteriori informazioni.

Consultare l'argomento "Connessione a un'origine dati" a pagina 14 per ulteriori informazioni.

Nota: I dati con caratteri non Latin non sono supportati. Nel caso in cui i dati siano formati da record in serie di caratteri Latin (ad esempio, Europa occidentale) e non Latin, verranno risolte solo le voci per i dati Latin.

Fase 2: creazione del 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.

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

Il processo di creazione di un repository è descritto in dettaglio più avanti. Consultare l'argomento "Impostazione di un repository di entità (nodo Esportazione EA)" a pagina 13 per ulteriori informazioni.

Nota: Se l'esecuzione è in modalità server remoto, è necessario creare il repository sulla macchina Modeler Server (cioè è necessario essere connessi a Modeler Server da Modeler Client quando si crea il repository in modo che il repository EA sia creato sulla macchina del server).

Una volta impostato un repository, è possibile gestirne il contenuto in vari modi. Consultare l'argomento "Configurazione di un repository di entità" a pagina 18 per ulteriori informazioni.

Fase 3: Connessione di SPSS Modeler al repository

Dopo avere creato il repository, lo si connette al flusso di SPSS Modeler.

Consultare l'argomento "Opzioni del repository di entità" a pagina 16 per ulteriori informazioni.

Fase 4: Mappatura dei 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 dati di input come campo **chiave univoco** che verrà utilizzato per le successive analisi. Inoltre, i campi dei dati di input devono essere mappati alle funzioni corrispondenti nel repository. La mappatura a funzionalità predefinite indica al repository entità quali campi mettere a confronto e, cosa più importante, come confrontarli. Il nodo Esportazione EA fornisce una tabella di mappatura in cui è possibile creare le mappature.

Consultare l'argomento "Mappatura dei campi di input alle funzioni (nodo Esportazione EA)" a pagina 16 per ulteriori informazioni.

Fase 5: esportazione dei dati nel repository e risoluzione delle identità

Ogni nodo origine dati richiede un proprio nodo Esportazione EA, pertanto se i dati sono distribuiti tra un numero diverso di origini, si potrebbe avere un flusso con diverse origini dati, ognuno dei quali è connesso ad un separato nodo Esportazione EA. Consultare l'argomento "Utilizzo di analisi delle entità con IBM SPSS Modeler" a pagina 5 per ulteriori informazioni.

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.

Si connette ogni nodo di origine dati al proprio nodo Esportazione EA, si mappano i campi di input alle funzioni del repository e quindi si esegue il flusso per esportare i dati da SPSS Modeler nel repository e

si risolvono i 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

Tabella 2. Dati esterni

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

Dipendenti

Tabella 3. Dipendenti

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

Clienti

Tabella 4. Clienti

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

Prestiti

Tabella 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 *\$EA-ID*) e un campo indicatore della sorgente (denominato *\$EA-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 6. Esempio di contenuto del repository dopo la fase di esportazione.

<i>\$EA-ID</i>	<i>\$EA-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	

Tabella 6. Esempio di contenuto del repository dopo la fase di esportazione (Continua).

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

Il sistema EA (Entity Analytics) ha determinato che *Mike* nel dataset *Esterno* è la stessa persona di *Michael* nel dataset *Dipendenti* basandosi sul numero di telefono che è in comune e gli ha assegnato l'ID 1.

Il caso di *Joe* nel dataset *Esterno* è un po' più complesso. È la stessa persona di nome *Joe* in *Clienti*? Non è possibile determinarlo sulla base delle sole due origini dati, ma abbiamo una terza origine dati *Prestiti*, contenente il nome *Joseph*. Esiste una corrispondenza: il numero di telefono di *Joseph* è lo stesso numero che appartiene a *Joe* nel dataset *Esterno*. In base a ciò, il sistema stabilisce che sono la stessa persona e attribuisce loro l'identificativo 2.

Non esistono più record per *Fred*, pertanto gli viene attribuito l'ID 3. *Susan* in *Clienti* viene identificata come la stessa persona di *Sue* in *Prestiti* poiché hanno lo stesso indirizzo, pertanto le viene attribuito l'ID 4.

Nota: questo è un esempio ottimale di corrispondenza 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 un'attività fraudolenta correlata all'esistenza di record duplicati per la stessa identità, si potrebbe voler produrre un report che elenca i duplicati.

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

L'output comune del nodo è costituito dai seguenti campi.

- Il campo identificativo aggiunto dal sistema (*\$EA-ID* nell'esempio Fase 5).
- Il campo indicatore origine aggiunto dal sistema (*\$EA-SRC* nell'esempio Fase 5)
- Il campo chiave univoco definito nella fase 4.

Inoltre, se si stanno esaminando le relazioni, viene prodotto il seguente output. Consultare l'argomento "Selezione di una sorgente dati" a pagina 26 per ulteriori informazioni.

- Il grado di separazione tra entità (*\$EA-DEGREE*)
- Il campo padre (*\$EA-PARENT*)
- Il campo figlio (*\$EA-CHILD*)
- La regola che identifica la relazione (*\$EA-RULE*)

Per visualizzare l'output in SPSS Modeler, è possibile collegare un nodo Output SPSS Modeler ad esempio un nodo Tabella o un nodo Report ed eseguire questa parte del flusso. Se è necessario riepilogare l'output che potrebbe essere molto grande, è possibile includere nodi Operazioni su record come ad esempio nodi Aggregazione o Selezione.

Il nodo origine Entity Analytics(EA) è descritto in dettaglio più avanti. Consultare l'argomento "Analisi delle identità risolte (nodo origine Entity Analytics(EA))" a pagina 26 per ulteriori informazioni.

Fase 7: Risoluzione di nuovi casi rispetto 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 verificare come si relazionano con ciò che si conosce per ottenere un miglior calcolo del punteggio? Qui interviene il nodo Streaming di EA.

Per prima cosa, si aggiunge un nuovo nodo origine dei dati di SPSS Modeler per leggere i nuovi dati nel flusso. Quindi, si connette il nodo origine a un nodo Streaming di EA. Per visualizzare l'output, si aggiunge un nodo Tabella come in precedenza.

Quando si esegue questa parte del flusso, il nodo Streaming di EA legge ogni nuovo record e lo confronta con il contenuto del repository. Se trova record corrispondenti nel repository, il nodo Streaming di EA indica nell'output tutti i record corrispondenti insieme al nuovo record a cui aggiunge i campi ID e indicatore origine. Se non trova corrispondenze, il nodo di elaborazione indica nell'output solo il nuovo record con i campi ID e indicatore origine aggiunti.

Per esempio, si supponga che il repository attualmente sia costituito dai contenuti generati dal nodo origine Entity Analytics(EA). Consultare Tabella 6 a pagina 8.

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

Tabella 7. 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 8. Output del nodo Streaming di EA.

\$EA-ID	\$EA-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
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 come chiave di aggregazione e passato ai nodi downstream per un'ulteriore elaborazione.

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

Fase 8: generazione 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 di business, 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 di business. Un'altra possibilità è esportare l'output in un database o qualche altro supporto per successive elaborazioni. IBM SPSS Modeler offre un'ampia gamma di azioni per soddisfare esigenze particolari.

Capitolo 3. Attività di analisi delle entità

Introduzione

Questa sezione descrive le seguenti attività di analisi 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
- Utilizzo delle analisi entità con gli altri prodotti IBM SPSS products
- Amministrazione dell'analisi entità

Impostazione di un repository di entità (nodo Esportazione EA)

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

1. Connessione a una sorgente dati Consultare l'argomento "Connessione a un'origine dati" a pagina 14 per ulteriori informazioni.
2. Creazione del repository. Consultare l'argomento "Creazione del repository" a pagina 14 per ulteriori informazioni.
3. Mappatura dei campi di input nella sorgente dati alle funzioni del repository. Consultare l'argomento "Mappatura dei campi di input alle funzioni (nodo Esportazione EA)" a pagina 16 per ulteriori informazioni.

Una volta impostate le mappature, è possibile visualizzarle per la sorgente dati corrente o per tutte le sorgenti dati conosciute nel repository. Consultare l'argomento "Visualizzazione delle mappature dei campi (nodo Esportazione EA)" a pagina 18 per ulteriori informazioni.

Nota: A partire dalla versione 16, SPSS Entity Analytics supporta i repository sul prodotto IBM DB2. Poiché un repository è specifico per una versione di SPSS Modeler e non è possibile importarlo da una versione precedente, se si dispone di un repository esistente e si sta eseguendo l'aggiornamento alla versione 16 di SPSS Entity Analytics sarà necessario ricreare tale repository nel nuovo database DB2.

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. Consultare l'argomento "Configurazione di un repository di entità" a pagina 18 per ulteriori informazioni.

Importante: dalla versione 16 a quelle successive IBM SPSS Modeler Entity Analytics supporta repository nel prodotto IBM DB2; le versioni precedenti di SPSS Entity Analytics supportavano repository in IBM solidDB. Se si dispone di un repository solidDB esistente, sarà necessario ricreare tale repository nel nuovo database DB2 se si effettuata l'aggiornamento a SPSS Entity Analytics versione 16 o successiva.

Nota: la versione di IBM SPSS Modeler Entity Analytics fornita con IBM SPSS Modeler Premium supporta solo un singolo repository ospitato nel prodotto IBM DB2 incluso con SPSS Entity Analytics. Con questa versione, è necessario eliminare un repository esistente prima di crearne uno nuovo. È disponibile un aggiornamento con licenza separata di SPSS 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 un'origine dati

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

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 origine all'area visualizzata.
2. Nell'area 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 Ricava. Consultare l'argomento "Aggiunta di un identificatore univoco del record" per ulteriori informazioni.

Nota: I dati con caratteri non Latin non sono supportati. Nel caso in cui i dati siano formati da record in serie di caratteri Latin (ad esempio, Europa occidentale) e non Latin, verranno risolte solo le voci per i dati Latin.

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 Ricava di SPSS Modeler.

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

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

Nota: Se l'esecuzione è in modalità server remoto, è necessario creare il repository sulla macchina Modeler Server (cioè è necessario essere connessi a Modeler Server da Modeler Client quando si crea il repository in modo che il repository EA sia creato sulla macchina del server).

Per creare un repository

1. Dalla scheda Esporta della palette dei nodi di SPSS Modeler, posizionare un nodo Esportazione EA nell'area del flusso.
Nota: se si crea un repository per la prima volta, utilizzare un nodo Esportazione EA e connetterlo al nodo origine SPSS Modeler contenente i dati che si desidera inserire nel repository (o il nodo Ricava se ne è stato aggiunto uno per ottenere un campo identificatore unico.). Per connettere i nodi, effettuare le seguenti operazioni.
 - a. Fare clic con il pulsante destro del mouse sul nodo origine di SPSS Modeler.
 - b. Scegliere Connetti.
 - c. 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

Fase 1

Qui si sceglie se creare un repository locale, mediante il prodotto IBM DB2 incluso con IBM SPSS Modeler Entity Analytics oppure se utilizzare un database esterno per il repository.

Crea repository locale. Specificare i dettagli per il nome utente e la password dell'amministratore per il database IBM DB2 che ospiterà il repository da creare. Confermare la password e fare clic su **Avanti**.

Nota: Non è possibile utilizzare un simbolo di tratteggio o di sottolineatura nel nome utente.

Le credenziali da utilizzare per il database IBM DB2 dipendono dal sistema operativo. Gli utenti UNIX devono utilizzare il nome utente g2user e la password G2password.

Le attività di amministrazione del repository all'interno dei nodi Entity Analytics, ad esempio la creazione o la distruzione di un repository, richiedono autorizzazioni aggiuntive. Su UNIX l'utente collegato a IBM SPSS Modeler Server deve essere l'utente root o g2user e un membro del gruppo db2iadm1. In Windows l'utente che ha eseguito l'accesso in IBM SPSS Modeler Server deve essere membro del gruppo DB2ADMNS per effettuare la gestione del repository.

se successivamente è necessario modificare le credenziali dell'amministratore, è possibile farlo utilizzando l'editor della riga di comando per il database. Consultare l'argomento "Gestione delle credenziali dell'amministratore per il database del repository" a pagina 33 per ulteriori informazioni.

Nota: Solo una combinazione di nome utente e password è possibile. Tutti gli utenti collegati al repository condividono lo stesso nome utente e password.

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**. Consultare l'argomento "Configurazione di un repository di entità" a pagina 18 per ulteriori informazioni.

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.

Connessione 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 un nome utente valido per il repository selezionato.
- **Password.** Password del nome utente.
- **Connetti.** Selezionare per connettersi al repository corrente.

Gestisci repository. 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 tutto.** 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. Consultare l'argomento "Svuotamento di un repository di entità" a pagina 35 per ulteriori informazioni.
- **Elimina non utilizzati.** Rimuove i dati di origine evidenziati dal repository ma conserva tutti i dettagli della configurazione. Consultare l'argomento "Eliminazione di origini dati non utilizzati da un repository" a pagina 35 per ulteriori informazioni.
- **Rinomina origine.** Apre una finestra di dialogo in cui è possibile modificare il nome dell'origine dati evidenziata.

Nota: Viene rinominata l'origine dati nel repository; sarà necessario selezionare di nuovo questo nuovo nome di origine dati in qualsiasi esportazione esistente o nodi di streaming da cui viene fatto riferimento.

Elimina intero repository. Cancella completamente i contenuti del repository corrente e i dettagli della configurazione. Consultare l'argomento "Eliminazione di un repository di entità" a pagina 36 per ulteriori informazioni.

Configura repository. Visualizza una finestra in cui è possibile configurare il repository corrente. Consultare l'argomento "Configurazione di un repository di entità" a pagina 18 per ulteriori informazioni.

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. Consultare l'argomento "Configurazione di un repository di entità" a pagina 18 per ulteriori informazioni.

Per mappare i campi di input alle funzioni

1. Collegare un nodo Esportazione EA a un nodo origine dei dati nell'area del flusso. Ogni nodo origine 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. Consultare l'argomento "Opzioni di input del repository per la mappatura" per ulteriori informazioni.
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 nodo di esportazione per un'origine dati differente), fare clic su **Esporta mappatura**.

Dopo avere mappato il primo nodo origine dei dati, ripetere la procedura per tutti gli altri nodi origine 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**.

Modalità. Mantenere la selezione di default, **Aggiungi al repository**, per aggiungere i record del file di origine 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 di 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. Consultare l'argomento "Opzioni del repository di entità" a pagina 16 per ulteriori informazioni.

Mappare a 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. Consultare l'argomento "Configurazione di un repository di entità" a pagina 18 per ulteriori informazioni.

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

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 alla 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. Consultare l'argomento "Configurazione di un repository di entità" per ulteriori informazioni.
- **Utilizzo.** Indica il contesto di un campo particolare quando sono possibili più contesti, ad esempio i numeri di telefono di casa e ufficio. Sono disponibili tipi di utilizzi predefiniti per le funzioni ADDRESS e PHONE ed è possibile creare propri tipi di utilizzi per tutte le funzioni. Per impostare un utilizzo diverso da quello predefinito (**Automatico**), fare clic su questa colonna nella riga che si desidera e scegliere uno dei tipi di utilizzo esistente (se esiste) oppure fare clic su <Aggiungi utilizzo...> per crearne uno nuovo. Consultare l'argomento "Gestione dei tipi di entità" a pagina 23 per ulteriori informazioni.

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 origine collegato a questo nodo di esportazione) o per tutte le sorgenti dati.

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.

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

<**Origine dati**>. In ogni colonna sono indicati i campi mappati in una determinata origine 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. Consultare l'argomento "Riutilizzo di una configurazione di repository" a pagina 25 per ulteriori informazioni.

Nota: A partire dalla versione 16, SPSS Entity Analytics supporta i repository sul prodotto IBM DB2. Poiché un repository è specifico per una versione di SPSS Modeler e non è possibile importarlo da una versione precedente, se si dispone di un repository esistente e si sta eseguendo l'aggiornamento alla

versione 16 di SPSS Entity Analytics sarà necessario ricreare tale repository nel nuovo database DB2.

Avvertenza:

Se si modifica e salva la configurazione di un repository che contiene dei dati, è possibile che venga richiesto di 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 9. Elementi principali della finestra Configurazione repository.

Sezione	Descrizione	
Sorgenti dati	Visualizza le mappature di tutte le sorgenti dati alle varie funzioni del repository.	Consultare l'argomento "Visualizzazione delle mappature delle sorgenti dati" per ulteriori informazioni.
Funzioni	Consente di creare una nuova funzione o di duplicare, modificare o eliminare una funzione già esistente.	Consultare l'argomento "Gestione delle funzioni del repository" a pagina 20 per ulteriori informazioni.
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).	Consultare l'argomento "Gestione dei tipi di entità" a pagina 23 per ulteriori informazioni.
Regole di risoluzione	Consente di impostare la soglia per la corrispondenza delle entità.	Consultare l'argomento "Impostazione della soglia per la corrispondenza delle entità" a pagina 25 per ulteriori informazioni.

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.

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

Nota: non è possibile aggiungere qui un'origine dati nel repository. Le origini dati possono essere aggiunte solo creando un nodo origine di SPSS Modeler e collegandolo a un nodo di esportazione di Entity Analytics. Consultare l'argomento "Connessione a un'origine dati" a pagina 14 per ulteriori informazioni.

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 attività verranno fornite più avanti in questa sezione.

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 o duplicate, ma possono essere solo salvate se si modifica l'attributo di anonimizzazione.

Frequenza. Indica quante identità possono avere lo stesso valore per questa funzione. I valori validi sono **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 deve generalmente avere solo uno di questi tipi di funzione. Ad esempio, una data di nascita o il numero della carta di identità qui deve assumere il valore **Sì** mentre un indirizzo o il numero della carta di credito deve avere il valore **No** (poiché un'entità potrebbe avere 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 è generalmente stabile durante l'intero ciclo di vita, ma poiché spesso viene specificato in modo non corretto, dando luogo a dati non corretti, la configurazione predefinita gli attribuisce il valore **No**.

Anonimizza. Indica se la funzione è stata anonimizzata. Le voci sono **Sì** o **No**. Consultare l'argomento "Anonimizzazione delle funzioni del repository" a pagina 22 per ulteriori informazioni.

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. Consultare l'argomento "Aggiunta o modifica di una funzione" a pagina 21 per ulteriori informazioni.

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**. Consultare l'argomento "Aggiunta o modifica di una funzione" per ulteriori informazioni.

Per modificare una funzione esistente

ATTENZIONE Se si modifica, elimina o si rende anonima una funzione o un elemento funzione quando il repository già contiene dati, si dovrebbe di conseguenza 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 quelle funzioni che sono state create, non le funzioni 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**. Consultare l'argomento "Aggiunta o modifica di una funzione" per ulteriori informazioni.

Per eliminare una funzione esistente

ATTENZIONE Se si modifica, elimina o si rende anonima una funzione o un elemento funzione quando il repository già contiene dati, si dovrebbe di conseguenza 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 quelle funzioni create, non le funzioni 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.

Avvertenza:

non è possibile annullare l'eliminazione di una funzione.

Aggiunta o modifica di una funzione

ATTENZIONE Se si modifica, elimina o si rende anonima una funzione o un elemento funzione quando il repository già contiene dati, si dovrebbe di conseguenza 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.

Nota: se una funzione esistente è bloccata, non è possibile modificarne i dettagli in questa finestra di dialogo.

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 **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 deve generalmente avere solo uno di questi tipi di funzione. Ad esempio, una data di nascita o il numero della carta di identità qui deve assumere il valore **Sì** mentre un indirizzo o il numero della carta di credito deve avere il valore **No** (poiché un'entità potrebbe avere 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 è generalmente stabile durante l'intero ciclo di vita, ma poiché spesso viene specificato in modo non corretto, dando luogo a dati non corretti, la configurazione predefinita gli attribuisce il valore **No**.

Tabella 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 può essere utilizzato per questo elemento. I tipi disponibili sono: Stringa, Intero, Reale e Data.

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

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

ATTENZIONE Se si modifica, elimina o si rende anonima una funzione o un elemento funzione quando il repository già contiene dati, si dovrebbe di conseguenza svuotare il repository e ricaricare i dati. In questo modo si evita di lasciare il repository in uno stato incoerente.

Anonimizza. Per la protezione dei dati è possibile scegliere di anonimizzare i dati quando vengono aggiunti a un repository; per abilitare questa opzione per una funzione, selezionare **Sì**. Consultare l'argomento "Anonimizzazione delle funzioni del repository" per ulteriori informazioni.

Anonimizzazione delle funzioni del repository

Come parte della sicurezza dei dati è possibile che si desideri anonimizzare i dati quando si aggiungono nel repository per ridurre il rischio che informazioni di identificazione personali possano essere inavvertitamente rivelate.

Quando i dati anonimizzati vengono esportati in un repository, è richiesto un metodo di anonimizzazione che consenta di intraprendere con i dati anonimi. Ad esempio, se due record di dati per i dettagli della carta di credito di una persona vengono anonimizzati come "anon_s21" e "anon_s9271", perdono la loro relazione; tuttavia, utilizzando un collegamento interno in background tra i record del sistema il sistema è ancora in grado di comprendere che un nome è un formato più breve dell'altro.

I collegamenti e gli identificativi in background che consentono di anonimizzare i dati da collegare vengono generati quando si crea un repository e sono univoci per il repository. I dati crittografati vengono archiviati internamente e quindi letti quando un flusso si collega a un repository.

Quando si configura il proprio repository, è possibile specificare se ogni singola funzione verrà anonimizzata. Se una funzione viene anonimizzata, tutti i relativi elementi vengono anonimizzati e viene sempre anonimizzata indipendentemente dal tipo di utilizzo. Consultare l'argomento "Aggiunta o modifica di una funzione" a pagina 21 per ulteriori informazioni.

Nota: Assicurarsi di non anonimizzare tutti i campi per SPSS Entity Analytics o non si sarà in grado di identificare quali dati vengono restituiti. Si consiglia di lasciare almeno un campo (anche se è solo un numero riga) senza anonimizzazione in modo da poter controllare successivamente la riunificazione ai dati originali.

Una colonna nell'elenco delle funzioni nella finestra Configurazione del repository mostra quali funzioni sono state impostate per l'anonimizzazione. Le voci sono **Sì** o **No**.

Nota: Se un repository esistente contiene dati prima che le funzioni vengano anonimizzate, si devono prima svuotare tutti i dati, altrimenti non vi saranno corrispondenze tra le funzioni anonimizzate e quelle non anonimizzate.

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

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. Consultare l'argomento "Opzioni di input del repository per la mappatura" a pagina 17 per ulteriori informazioni.

Informazioni generali relative ai tipi di utilizzo:

- I tipi di utilizzo sono etichette arbitrarie.
- È possibile creare un tipo di utilizzo da un'immissione di testo recente; tuttavia non è consentito immettere spazi e caratteri non validi.
- Ciò che viene immesso, viene automaticamente modificato in maiuscolo al momento dell'immissione.
- È possibile avere tanti tipi di utilizzo quanti se ne desiderano.
- I tipi di utilizzo non devono essere significativi, ma è utile quando si esegue successivamente il mapping se si utilizza una convenzione di denominazione che abbia senso per l'utente e per gli altri utenti.
- Quando si esegue il mapping, viene visualizzato un avviso in un tipo di carattere rosso se si sono mappati alcuni elementi con un tipo di utilizzo e altri con un differente tipo.

Generalmente viene visualizzato un errore se si tenta di mappare due campi allo stesso elemento funzione. I tipi di utilizzo rappresentano un criterio per mappare due o più campi allo stesso elemento funzione ed essere in grado di creare corrispondenze tra di loro.

Ad esempio, se si sono definite due funzioni separate: *HOMEADDRESS* e *WORKADDRESS*, non ci dovrebbe essere nessuna corrispondenza tra di loro. Se un'entità ha un *HOMEADDRESS* che è lo stesso dell'altra entità *WORKADDRESS* non esistono corrispondenze perché sono differenti funzioni. Tuttavia, se si riutilizza una singola funzione con differenti tipi di utilizzo, la risoluzione riconosce che *ADDRESS.WORK* è lo stesso di *ADDRESS.HOME*.

È possibile riutilizzare tipi di utilizzo per differenti funzioni o averne diversi; ad esempio *HM* e *WK* per il telefono e *HOME* e *WORK* per l'indirizzo. Non ha importanza perché non esiste corrispondenza tra i telefoni e gli indirizzi; tuttavia, essere congruenti aiuterà l'utente ad identificare e raggruppare i campi successivamente.

Quando più tipi di entità sono immessi in un singolo repository, a patto che l'utente stia utilizzando la stessa funzione, non ha importanza quali tipi di utilizzo sono. Ad esempio, se si definisce *WK* e *HM* come tipi di utilizzo di *ADDRESS* per il tipo di entità *COMPANY* esisterà ancora corrispondenza tra *WORK* e *HOME* in quanto tipi di utilizzo di *ADDRESS* per *PERSON*.

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, elimina o si rende anonima una funzione o un elemento funzione quando il repository già contiene dati, si dovrebbe di conseguenza 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, elimina o si rende anonima una funzione o un elemento funzione quando il repository già contiene dati, si dovrebbe di conseguenza 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.

Avvertenza:

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.

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

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

Scegliere **Imposta per risoluzione predefinita** per tornare alla soglia predefinita da una delle altre impostazioni.

Scegliere **Imposta per risoluzione conservativa** se vengono trovate troppe corrispondenze.

Per compilare un repository per entrambe le entità e le relazioni, selezionare **Includi relazioni**. Si noti che questa opzione è solo disponibile se si dispone dell'aggiornamento con licenza separata noto come IBM SPSS Modeler Entity Analytics Unleashed.

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). Questa funzione è possibile solo in una configurazione esistente. Ad esempio, è impossibile migrare una configurazione di repository da una versione di IBM SPSS Modeler a un'altra o da un tipo di database a un altro.

Per riutilizzare una configurazione esistente

1. Visualizzare la finestra Configurazione repository per il repository di cui si desidera utilizzare la configurazione. Consultare l'argomento "Configurazione di un repository di entità" a pagina 18 per ulteriori informazioni
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 origine Entity Analytics(EA))

Una volta esportati i dati nel repository, è possibile utilizzare il nodo origine 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 origine Entity Analytics(EA) a un flusso.
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). Consultare l'argomento "Selezione di una sorgente dati" per ulteriori informazioni
4. Aggiungere ulteriori nodi al flusso per eseguire l'elaborazione desiderata. Consultare l'argomento "Aggiunta di nodi al flusso" a pagina 28 per ulteriori informazioni.

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 origini dati elencate.

Repository di entità. Visualizza il repository di entità corrente, se esiste. Per scegliere un repository diverso nel caso in cui ne esista più di uno, selezionarne uno dall'elenco. Per creare un nuovo repository, scegliere <Cerca...> per visualizzare una finestra di dialogo da cui è possibile creare il repository. Consultare l'argomento "Opzioni del repository di entità" a pagina 16 per ulteriori informazioni.

Includi record dalle origini dati. In questa tabella sono elencate le diverse origini dati che sono state inserite nel repository, unitamente al numero di record in ogni sorgente. Selezionare la casella di controllo **Includi** per le origini dati che si desidera utilizzare per ulteriori analisi ed elaborazioni. Per selezionare o deselezionare tutte le origini dati, fare clic su **Includi tutto** o **Escludi tutto** rispettivamente.

Relazioni. Selezionare il tipo di relazioni da includere nel repository. Si noti che questa opzione è solo disponibile se si dispone dell'aggiornamento con licenza separata noto come IBM SPSS Modeler Entity Analytics Unleashed.

- **Nessuna relazione.** I dettagli delle relazioni non vengono utilizzati.
- **Relazioni strette.** Seleziona solo le entità strettamente correlate. Il criterio di vicinanza di una relazione dipende da molte variabili, come le proprietà delle funzioni mappate, quali funzioni sono condivise e se la risoluzione è impostata per essere conservativa o aggressiva.
- **Tutte le relazioni.** Seleziona tutte le entità correlate.

Max. grado di separazione. Disponibile solo se sono selezionate le opzioni **Relazioni strette** o **Tutte le relazioni**. Selezionare il numero di gradi di separazione da utilizzare per identificare una relazione. Ad esempio, se Anna e Roberto non si conoscono, ma Giovanni conosce entrambi, Anna e Roberto sono correlati da due gradi di separazione.

Tipo di entità di output- Per impostazione predefinita, se il repository contiene dettagli, viene mostrato il primo tipo di identità elencato nel repository. Se il repository ne possiede più di uno, selezionando il tipo di entità qui presente si modificano le funzioni mostrate nella scheda Filtro per elencare le funzioni per quel tipo. È possibile selezionare un qualsiasi tipo di entità utilizzato nel repository.

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 10. Campi identità risolte

Campo	Descrizione
\$EA-ID	Identificatore entità
\$EA-SRC	Tag origine 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: sebbene sia possibile utilizzare la scheda Filtri per escludere i campi, non si dovrebbe fare poiché i campi di identità risolte rappresentano il minimo necessario per elaborare le analisi dell'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 11. 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.
Verifica	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 nugget del modello.

Aggiunta di nodi al flusso

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

- Nodo Aggregazione o Elimina duplicati 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 origine 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 origine, 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 origine 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. Consultare l'argomento "Output del nodo Streaming di EA" a pagina 31 per ulteriori informazioni.

Per confrontare i nuovi casi in base al repository

1. Connettersi alla sorgente dati contenente i nuovi record da confrontare con le entità esistenti. Consultare l'argomento "Connessione a un'origine dati" a pagina 14 per ulteriori informazioni.
2. Dalla scheda Oper su record, collegare un nodo Streaming di EA al nodo origine 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 mappare. Consultare l'argomento "Gestione dei tipi di entità" a pagina 23 per ulteriori informazioni.
10. Mappare i campi di input nella sorgente dati alle funzioni del repository. Consultare l'argomento "Mappatura dei campi di input alle funzioni (nodo Streaming di EA)" a pagina 29 per ulteriori informazioni.
11. Facoltativamente, se si desidera aggiornare i record nel repository in tempo reale mentre si calcolano i dati. Consultare l'argomento "Mappatura dei campi di input alle funzioni (nodo Streaming di EA)" a pagina 29 per ulteriori informazioni.

12. Fare clic sulla scheda **Output** per visualizzare i dettagli delle diverse origini dati inserite nel repository e impostare i criteri di selezione per recuperare le entità esistenti. Consultare l'argomento "Visualizzazione delle mappature dei campi e delle origini dati (nodo Streaming di EA)" a pagina 30 per ulteriori informazioni.
13. Fare clic sulla scheda **Filtro** per visualizzare i dettagli dei campi di input e le funzioni archiviate nel repository. Tutte le funzioni che non sono state mappate nel nodo sono filtrate per impostazione predefinita; tuttavia è possibile effettuare modifiche se richiesto.
14. Fare clic su **OK** quando il nodo è impostato correttamente.
15. Collegare un nodo Tabella al nodo Streaming di EA ed eseguire il flusso.

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-**. Consultare l'argomento "Output del nodo Streaming di EA" a pagina 31 per ulteriori informazioni.

Nota: si potrebbe verificare un errore del tipo **Numero di campi non corretto rilevati nel modello dati del server** quando si esegue il nodo Streaming di EA. Ciò accade se si è modificata la configurazione del repository a partire dalla creazione del nodo Streaming di EA. La modifica della configurazione in questi casi può avere l'effetto di modificare il numero e i nomi dei campi emessi dal nodo. Per risolvere il problema, aprire il nodo Streaming di EA e fare clic sul pulsante **Aggiorna**. In questo modo il numero e i nomi dei campi di output verrà ricalcolato.

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 di mappatura in questa scheda o selezionare la scheda **Visualizza** per visualizzare i dettagli di tutte le origini 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**.

Repository di 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. Consultare l'argomento "Opzioni del repository di entità" a pagina 16 per ulteriori informazioni.

Mappare a 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. Consultare l'argomento "Configurazione di un repository di entità" a pagina 18 per ulteriori informazioni.

Mantieni ricerche. Se si desidera aggiornare i record nel repository in tempo reale mentre si calcolano i dati, selezionare questa opzione.

Tag origine. Disponibile solo quando si seleziona **Mantieni ricerche**. Un elenco di tag che indica le origini dati attualmente note al repository. Scegliere una voce dall'elenco oppure <**Aggiungi nuovo tag sorgente...**> per creare un tag per una nuova sorgente dati.

Chiave univoca. Disponibile solo quando si seleziona **Mantieni ricerche**. 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 alla 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. Consultare l'argomento "Configurazione di un repository di entità" a pagina 18 per ulteriori informazioni.
- **Utilizzo.** Indica il contesto di un campo particolare quando sono possibili più contesti, ad esempio i numeri di telefono di casa e ufficio. Consultare l'argomento "Gestione dei tipi di entità" a pagina 23 per ulteriori informazioni.

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 e delle origini dati (nodo Streaming di EA)

Nella scheda Output è possibile vedere i dettagli delle diverse origini 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.

Includi corrispondenze dalle origini dati. In questa tabella sono elencate le diverse origini dati che sono state inserite nel repository, unitamente al numero di record in ogni origine.

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.

Nota: se vengono rilevate più di 20 corrispondenze, verranno restituite solo le prime 20 trovate.

- **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 quando si desidera restituire solo quelle entità che sono considerate corrispondenze esatte.
- **Includi corrispondenze possibili.** Utilizzare questa impostazione quando si desidera restituire sia le entità corrispondenti che le entità che condividono gli stessi identificativi (quelle con funzioni configurate con il valore di frequenza Uno, ad esempio, numeri di carte di credito corrispondenti, codici fiscali e così via).
- **Includi tutte le corrispondenze.** Utilizzare questa opzione quando si desidera visualizzare nel repository il maggior numero possibile di entità che hanno funzioni condivise. Si tratta del criterio di corrispondenza meno severo e dà come risultato il maggior numero di record selezionati. Questa opzione restituisce le corrispondenze esatte e entità che condividono quasi qualsiasi funzione (generalmente quelle con un valore di frequenza Uno o Pochi). Ad esempio, sarebbero incluse entrambe le entità con lo stesso codice fiscale ed entità con indirizzi simili.

Relazioni. Disponibile solo se il repository è stato configurato per includere le relazioni. Per configurare il repository in modo da includere le relazioni, si deve disporre dell'aggiornamento con licenza separata come IBM SPSS Modeler Entity Analytics Unleashed. Selezionare il tipo di relazioni da includere nell'output.

- **Nessuna relazione.** I dettagli delle relazioni non vengono utilizzati.
- **Relazioni strette.** Seleziona solo le entità strettamente correlate. La vicinanza di una relazione dipende da molte variabili, ad esempio le proprietà delle funzioni mappate, quali funzioni vengono condivise e se la risoluzione viene impostata per essere conservativa o aggressiva.

- **Tutte le relazioni.** Seleziona tutte le entità correlate.

Max. grado di separazione. Disponibile solo se sono selezionate le opzioni **Relazioni strette** o **Tutte le relazioni**. Selezionare il numero di gradi di separazione da utilizzare per identificare una relazione. Ad esempio, se Anna e Roberto non si conoscono, ma Giovanni conosce entrambi, Anna e Roberto sono correlati da due gradi di separazione.

Tipo di entità di output- Per impostazione predefinita, se il repository contiene dettagli, viene mostrato il primo tipo di identità elencato nel repository. Se il repository ne possiede più di uno, selezionando il tipo di entità qui presente si modificano le funzioni mostrate nella scheda Filtro per elencare le funzioni per quel tipo. È possibile selezionare un qualsiasi tipo di entità utilizzato nel repository.

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[, Campo2[, ... FieldN]]</i>	Campi della sorgente dati che contiene i nuovi record.
\$EA-ID	Identificatore di entità per questo record nel repository.
\$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-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-Funzione1[, \$EA-Funzione2[, ... \$EA-FeatureN]]	Valori delle funzioni mappate per questo record nel repository.

Se i campi delle relazioni sono abilitati nel repository e il grado di separazione è superiore a zero nella scheda Output, l'output del nodo Streaming di EA è costituito anche dai seguenti campi per ogni record recuperato.

Campo	Descrizione
\$EA-DEGREE	Grado di separazione.
\$EA-PARENT	Identificatore del record da cui viene calcolata la separazione.
\$EA-CHILD	Identificatore del record rispetto al quale viene calcolata la separazione.
\$EA-RULE	

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*.

Dopo l'installazione, bisogna utilizzare il client IBM SPSS Collaboration and Deployment Services Deployment Manager per creare una definizione di server del repository di Entity Analytics. Questa è richiesta per l'utilizzo di un flusso IBM SPSS Modeler che contiene un nodo Entity Analytics in un job IBM SPSS Collaboration and Deployment Services (in altre parole, per eseguire i flussi Entity Analytics in IBM SPSS Collaboration and Deployment Services). La definizione del server deve corrispondere con il nome del repository nel flusso: questa definizione viene utilizzata per comunicare al flusso dove trovare il repository e fornirgli le informazioni di connessione di cui ha bisogno.

Attività amministrative

Per i repository creati con Entity Analytics, viene creato un nuovo servizio database con il prodotto IBM DB2. Vi sono alcune attività amministrative associate a DB2, che in genere vengono eseguite dall'amministratore di 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 del repository in una differente directory di archiviazione
- Impostazioni delle proprietà flusso per i campi data/ora e timestamp.
- Modifica delle impostazioni di timeout
- 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 DB2 venga allocata una porta che non può essere allocata 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 IBM 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 installazione Modeler Server >/ext/bin/pasw.entityanalytics/ea.cfg` e impostando i valori opportuni per le impostazioni `min_port` e `max_port`. Il contenuto di default di questo file è illustrato di seguito.

```
# port range configuration for entity analytics
#
#   this port range controls which ports DB2 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 DB2 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 DB2.

Per avviare l'editor SQL di DB2

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

```
cd modeler_install_dir\ext\bin\pasw.entityanalytics\DB2\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\version\EA\repositories\repos_name"
```

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 dell'amministratore del database corrente per visualizzare il prompt `solsql>`.

Per modificare la password dell'amministratore di database

1. Al prompt `solsql>` immettere:

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

dove *nome utente* è il nome utente corrente dell'amministratore di database e *password* è la nuova password.
2. Immettere Uscita oppure chiudere l'editor.
3. Riavviare il client SPSS Modeler.

Per informazioni sulle altre attività amministrative da eseguire con il database DB2, consultare la documentazione relativa alla versione appropriata di IBM DB2 all'indirizzo <http://publib.boulder.ibm.com/>.

Spostamento del repository ad una directory di archiviazione diversa

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.

Per spostare il repository in una diversa directory

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

Impostazioni delle proprietà flusso per i campi data/ora e timestamp.

Se i dati di origine includono campi contenenti dati data/ora o timestamp, accertarsi che le corrispondenti proprietà flusso siano impostate nel formato riconosciuto da IBM SPSS Modeler Entity Analytics.

Per impostare il formato della proprietà flusso

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**.

Modifica delle impostazioni di timeout

Sui sistemi lenti o sovraccarichi, se si rilevano degli errori durante la creazione o l'accesso ai repository è possibile che sia necessario incrementare le impostazioni di timeout per avviare ed arrestare il motore dell'analisi entità o il server di database dell'analisi entità.

Per modificare il timeout per il motore delle analisi entità

1. Uscire da SPSS Modeler.
2. Modificare il file *<percorso di installazione server modeler>/ext/bin/pasw.entityanalytics/ea.cfg* per incrementare il valore della seguente opzione:
timeout, *value*

in cui *valore* è il valore di timeout espresso in secondi per il motore delle analisi entità (il valore predefinito è 60).

Per modificare il timeout per il server di database di analisi delle entità (solo DB2)

1. Uscire da SPSS Modeler.
2. Modificare il file *<percorso di installazione server modeler>/ext/bin/pasw.entityanalytics/ea.cfg* per incrementare il valore della seguente opzione:
timeout, *value*

dove *valore* è il valore di timeout espresso in secondi per il server di database DB2 di analisi entità (il valore predefinito è 100).

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: in particolare se si utilizza un client SPSS Modeler a 32 bit e un SPSS Modeler Server a 64 bit (o viceversa sarà necessario seguire questa procedura.

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* per utilizzare porte diverse da quelle dell'altro sistema. Consultare l'argomento "Configurazione delle assegnazioni delle porte" a pagina 32 per ulteriori informazioni.
3. Modificare l'impostazione *repository_data_directory* per utilizzare una directory diversa da quelle dell'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, è possibile eliminare i dati dal repository.

Per eliminare tutti i dati da 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. Fare clic sul pulsante **Elimina tutto** quando lo si abilita.
8. Nella finestra di dialogo Elimina tutte le origini dati, fare clic su **Elimina** per confermare l'eliminazione del repository.

Eliminazione di origini dati non utilizzati da un repository

Se si dispone di un'origine dati che non si utilizza più o non è più necessaria in un repository entità, è possibile eliminare l'origine dal repository. È possibile selezionare una o più origini dati da eliminare.

Per eliminare un'origine dati da 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 un repository da cui si desidera eliminare un'origine dati.
6. Se non è già stata eseguita la connessione, immettere nome utente e password dell'amministratore e fare clic su **Connetti**.
7. Nell'elenco **Gestisci repository** selezionare l'origine dati da eliminare. Se necessario, fare clic tenendo premuto il tasto Ctrl per selezionare ulteriori origini dati.
8. Quando il pulsante **Elimina non utilizzati** è abilitato, fare clic su di esso.

9. Nella finestra di dialogo Elimina origini dati non utilizzate, fare clic su **Elimina** per confermare l'eliminazione del repository.

Eliminazione di un repository di entità

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

Attenzione: questo è esattamente ciò che si dice. **Non è possibile annullare questa operazione.** 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. Consultare l'argomento "Svuotamento di un repository di entità" a pagina 35 per ulteriori informazioni.

Nota: la seguente procedura presuppone che sia possibile connettersi al repository da SPSS Modeler, e che si conosca il nome utente e la password dell'amministratore per il database che ospita il repository. In caso contrario, attenersi alla procedura di eliminazione di un repository quando non è possibile connettersi a esso. Consultare l'argomento "Eliminazione di un repository quando non è possibile connettersi a esso" per ulteriori informazioni.

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

Utilizzare la 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 questa procedura sul computer che ospita il database del repository.

Sistemi Windows

1. Aprire una finestra di richiesta comandi.
2. Immettere:

```
cd modeler_install_dir
cd ext\bin\pasw.entityanalytics
delete_repository.bat repos_name
```

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

Nota: Il nome file del repository è sensibile al maiuscolo/minuscolo

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

Sistemi UNIX

1. Aprire una shell.
2. Immettere:

```
cd modeler_server_install_dir  
cd ext/bin/pasw.entityanalytics  
./delete_repository.sh repos_name
```

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

Nota: Il nome file del repository è sensibile al maiuscolo/minuscolo.

Completamento della procedura (tutti i sistemi)

1. Al prompt, confermare l'eliminazione del repository digitando Y.
2. Eliminare la directory con lo stesso nome del repository eliminato. Se non è possibile eliminare la directory, riavviare il computer e riprovare.

Capitolo 4. Funzionamento di Entity Analytics

Informazioni su questo esempio

In questo esempio, verrà descritto come l'aggiunta di analisi entità può migliorare ulteriormente i già straordinari risultati che è possibile ottenere dall'utilizzo di IBM SPSS Modeler.

Questo esempio utilizza il flusso 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 IBM SPSS Modeler in cui è installato anche IBM SPSS Modeler Entity Analytics. È possibile accedere alla directory *Demos* dal gruppo di programmi IBM 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 sistema. Provvedere subito prima di continuare con questo esempio. Consultare l'argomento "Creazione del repository" a pagina 14 per ulteriori informazioni.

Incominciamo con un caso familiare – i dirigenti di una banca sono preoccupati della possibilità che i clienti possano risultare inadempienti rispetto ai prestiti ottenuti per cui le richieste vengono sospese. Il reparto IT della banca utilizza da molto tempo SPSS Modeler, quindi ha già creato un flusso 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 propri modelli e ciò che da essi hanno appreso.

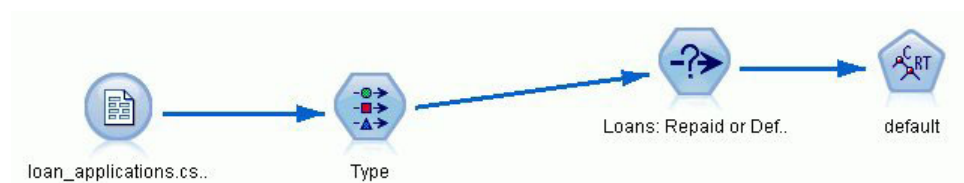


Figura 2. Flusso iniziale con nodo Modelli

Oltre ai dettagli dei prestiti concessi in passato, il dataset *loan_applications.csv* include i dettagli di 150 clienti le cui richieste di prestito sono ancora in corso per un totale di 850 record.

Non tutti i campi del dataset sono utili per effettuare previsioni, ad esempio, i campi nome 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 3. Ruoli dei campi impostati nel nodo Tipo

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

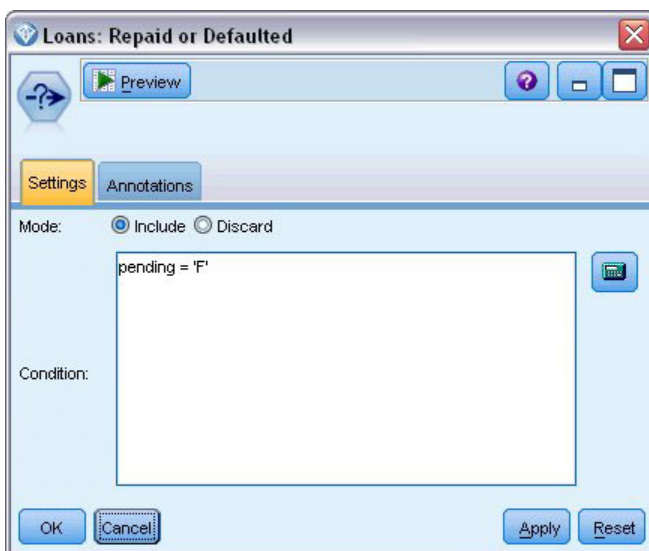


Figura 4. Eliminazione delle richieste di prestito in sospeso

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, si è utilizzato un nodo C&R Tree

che verrà utilizzato per creare un modello per prevedere i potenziali clienti insolventi sulla base delle prestazioni passate dei clienti della banca.

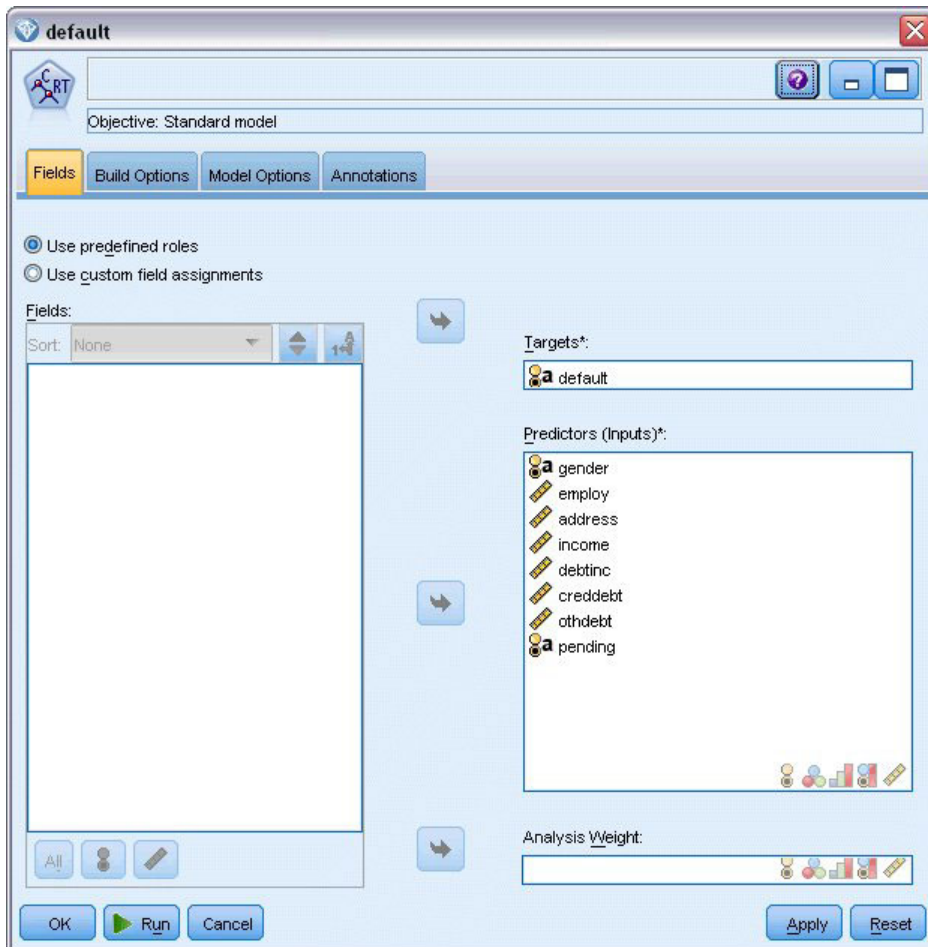


Figura 5. Assegnazione di campi predittori e obiettivi

I campi utilizzati per effettuare la previsione vengono indicati come campi predittori e il campo il cui valore il modello sta tentando di prevedere **default** in questo caso è impostato come campo obiettivo come definito in precedenza dal nodo Tipo.

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

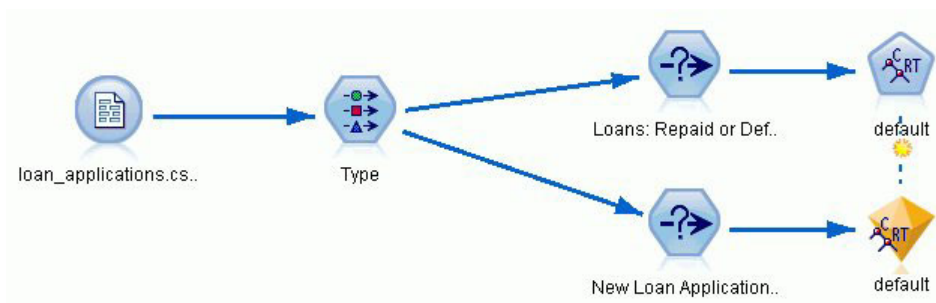


Figura 6. Flusso con aggiunta del nugget del modello

Ora l'analista della banca può utilizzare il modello per avviare la previsione se i clienti con rimborsi in sospeso probabilmente saranno insolventi. Utilizzando il dataset originale, l'analista inserisce un nodo Selezione che questa volta include solo i 150 record di prestiti contrassegnati come In sospeso invece di eliminarli. L'analista trasferisce questi record direttamente nel modello aggiungendo un nodo Distribuzione per una rappresentazione visiva delle previsioni del modello.

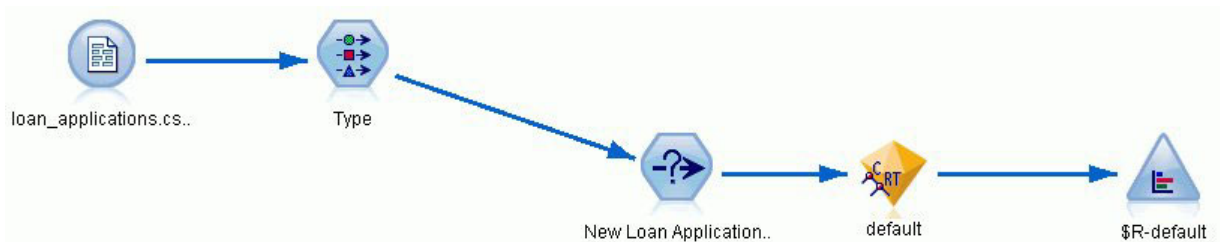


Figura 7. Flusso con la selezione delle nuove richieste di prestito e l'aggiunta del nodo distribuzione

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 è 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 del flusso, l'analista apprende dall'output del nodo Distribuzione che 137 dei 150 nuovi richiedenti rimborseranno i prestiti. I rimanenti 13 probabilmente saranno inadempienti, quindi l'analista consiglierà alla banca di rifiutare le loro richieste.

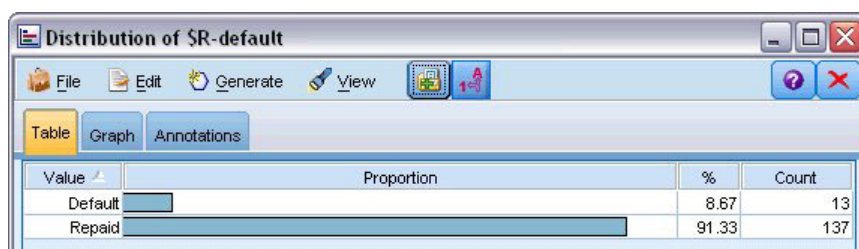


Figura 8. Output del nodo distribuzione senza analisi delle entità

Aggiunta dell'analisi delle entità

Ora verifichiamo se la situazione può essere migliorata aggiungendo analisi 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. Potrebbero esserci record duplicati causati da errori nell'immissione dati, ma è anche possibile che un richiedente un mutuo abbia tentato di camuffare la propria identità. In ogni caso, la banca deve avere il quadro corretto della situazione.

A scopo di esempio, si presuppone che un repository entità è già stato creato. Consultare l'argomento "Creazione del repository" a pagina 14 per ulteriori informazioni.

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.

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 (note come "funzioni") per evitare duplicazioni.

Nel nodo Esportazione EA, impostare i dettagli relativi al repository: i dettagli della connessione, il tag origine (per identificare l'origine dati – **TEST** in questo caso), il tipo di entità (l'insieme di funzioni che si utilizza – quello denominato **PERSON**) e il campo chiave univoco (per identificare univocamente ogni record). In questo caso, utilizzare il campo **key** come chiave univoca.

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

Iniziare impostando la 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 rimanenti campi che corrispondono alle funzioni in modo che l'intero insieme di mappature sia simile a quello riportato di seguito.

Tabella 12. 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

Una volta esportati i dati nel repository, il sistema di analisi entità avvia la risoluzione di possibili conflitti di identità assegnando un identificativo entità univoco che verrà visualizzato successivamente nel campo *\$EA-ID*. (Nota: questo non è lo stesso campo Chiave univoca nel nodo Esportazione EA che viene utilizzato per identificare univocamente i record dell'origine dati.)

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

Aprire il nodo origine Entity Analytics(EA) e impostare i dettagli del repository di entità. Viene quindi visualizzato un elenco delle origini dati che sono state esportate nel repository in questo caso ne esiste solo una.

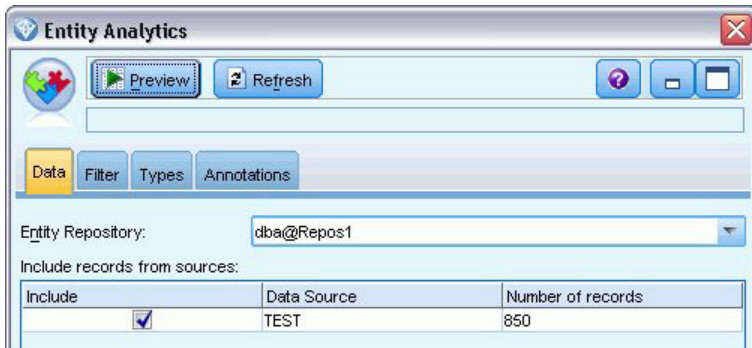


Figura 9. Selezione di un'origine dati nel repository

Selezionare la casella di controllo per l'origine dati **TEST** e fare clic su OK.

Osserviamo cosa ha fatto il sistema analisi entità sui dati. Collegare un nodo Tabella al nodo origine Entity Analytics(EA), aprire il nodo Tabella e fare clic su **Esegui** per visualizzare la finestra di 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

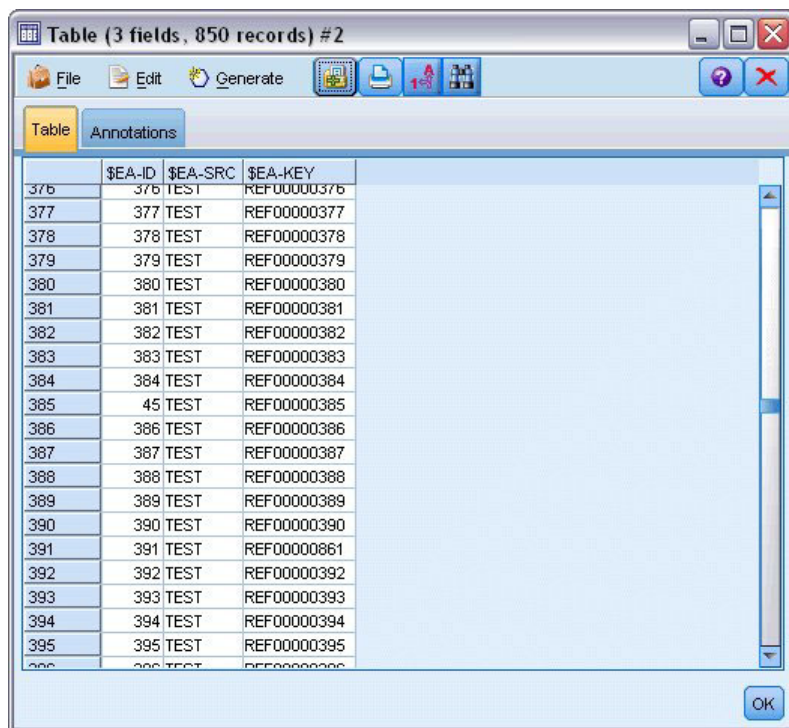
Figura 10. Output del nodo Tabella

Solo un campo risulta familiare- quello etichettato *\$EA-KEY*. Questo è effettivamente il campo *chiave* dell'origine dati ed è presente qui perché è scelto come campo Chiave univoca nel nodo Esportazione EA.

Il sistema ha aggiunto altri due campi. Il campo *\$EA-ID* rappresenta l'identificativo univoco, non i record origine, ma delle identità risolte. Tra poco vedremo le differenze. Il campo *\$EA-SRC* identifica da dove provengono i dati in questo caso **TEST** perché questo è il tag origine assegnato al campo nel nodo Esportazione EA.

Cosa è accaduto a tutti gli altri campi nei dati di origine? Non preoccuparti, sono ancora nel repository - è solo per motivi di prestazioni che il nodo origine EA (Entity Analytics) passa solo il minimo insieme di campi a valle per ulteriori elaborazioni.

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



	\$EA-ID	\$EA-SRC	\$EA-KEY
376	376	TEST	REF00000376
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

Figura 11. Differenze tra le righe di output di Tabella e i numeri di \$EA-ID

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 REF0000045. Scorrendo verso la fine l'output, si trovano più numeri fuori sequenza, ad esempio, alle righe 485, 517, 520 e così via. Dovremmo approfondire.

Prima evidenziamo il fatto che il campo *\$EA-KEY* contiene i dati del campo *chiave* nei dati di origine ridenominandolo in *chiave*. Collegare un nodo Filtro al nodo origine Entity Analytics(EA) e aprire il nodo Filtro. Fare doppio clic sulla stringa *\$EA-KEY* nella seconda colonna **Campo** e immettere *key*.

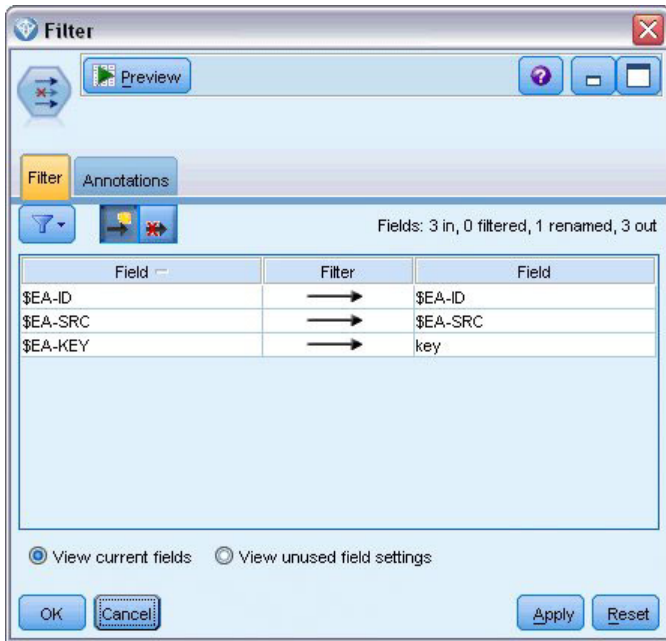


Figura 12. Ridenominazione del campo \$EA-KEY

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 13. Ordinamento degli ID entità in ordine crescente

Lasciare selezionato il criterio di ordinamento **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 Ricava al nodo Ordina. Aprire il nodo Ricava e impostare il nome **Ricava** su IsDuplicate. Dall'elenco **Deriva come**, scegliere **Flag**, che imposta anche il **Tipo di campo** su **Flag**. Impostare il campo **Valore vero** su **Duplicate** e il campo **Valore falso** su **Unique**.

Per trovare i record duplicati, si farà uso di una funzione di sequenza speciale denominata @OFFSET, che viene fornita con SPSS Modeler.

Immettere quanto segue nel campo **If**:

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



Figura 14. Impostazione della condizione nel nodo Ricava

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 Ricava, allegare un nodo Tabella al nodo Ricava, 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.

	\$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

Figura 15. Output del nodo Ricava

Ricordi quando abbiamo osservato l'output direttamente dal nodo origine EA (Entity Analytics)? Il sistema ha già identificato che il record REF00000385 fa riferimento alla stessa persona dell'entità 45. Ora si è eseguita un'ulteriore fase e indicato il fatto che i record REF00000045 e REF00000385 sono duplicati poiché entrambi fanno 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 Ricava *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.

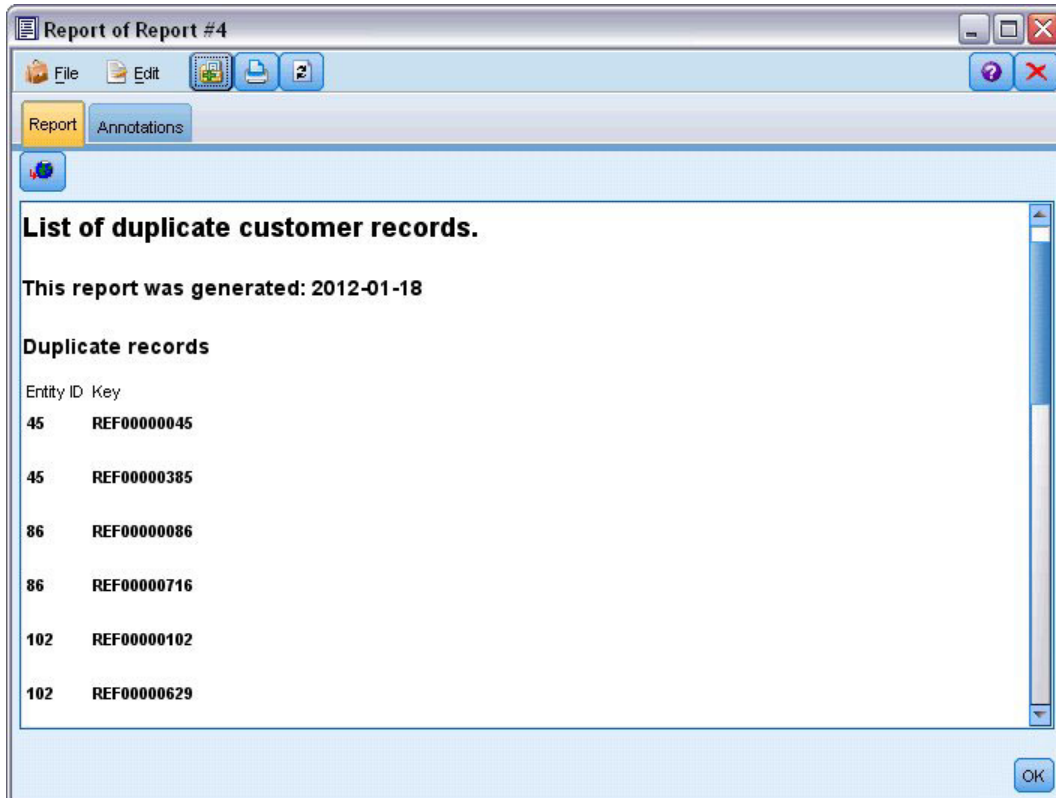


Figura 16. Output del nodo report

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 di questo esempio è verificare se l'aggiunta di analisi delle entità produce qualche differenza rispetto alla previsione originale della banca. Come ricordiamo il modello originale aveva previsto 13 inadempienti delle 150 richieste di prestito ancora in corso. Si è in procinto di utilizzare un nodo Unione per unire l'output del modello con le informazioni relative ai record duplicati delle analisi delle entità per esaminare se in questo caso la previsione cambia.

Prima, è necessario garantire che i nuovi campi aggiunti mediante le analisi entità hanno il corretto tipo di dati o *livelli di misurazione* come sono conosciuti in SPSS Modeler. Collegare un nodo Tipo al nodo Ricava **IsDuplicate**, aprire il nodo Tipo e fare clic sul pulsante **Leggi valori**.

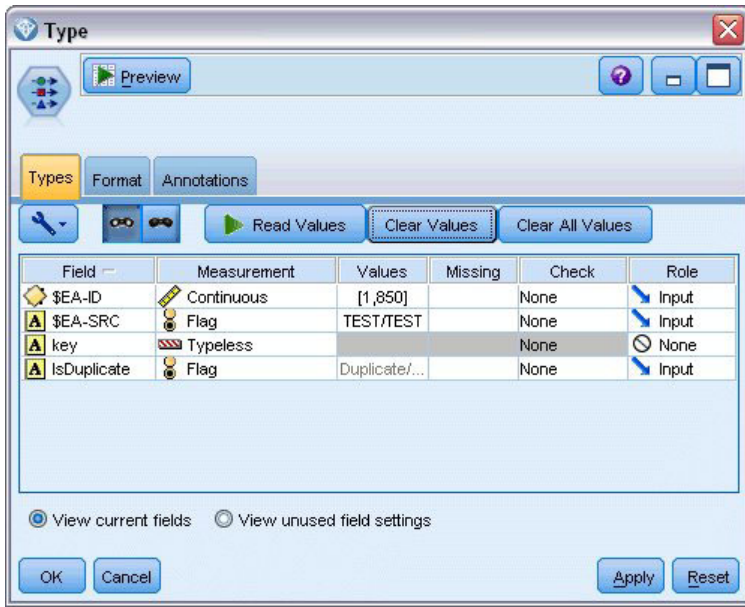


Figura 17. Impostazioni per il nodo Tipo

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

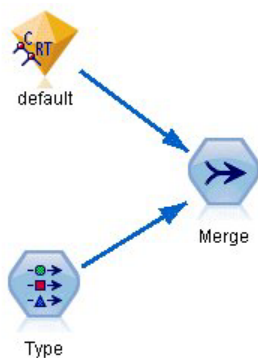


Figura 18. Input per il nodo Unione

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

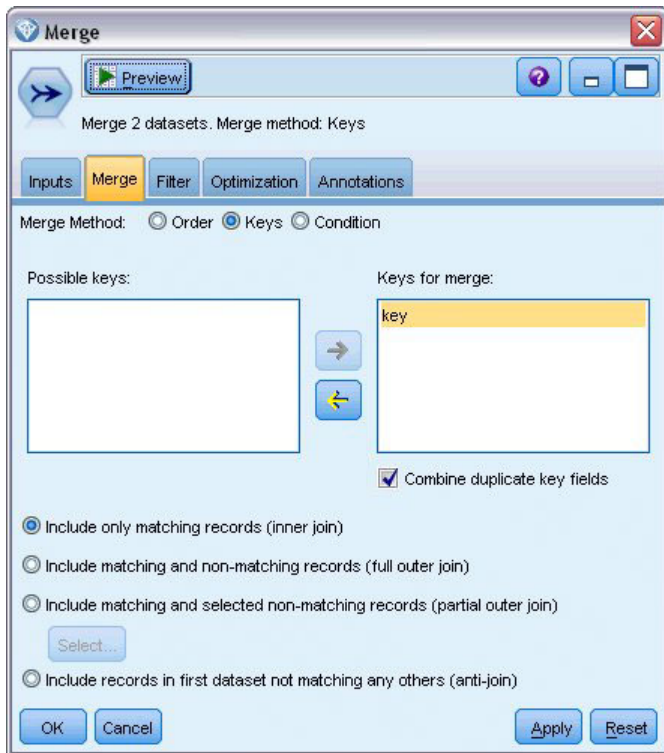


Figura 19. Specifica del campo chiave per l'operazione di unione

Ora si è quasi pronti ad eseguire il confronto. Tuttavia, se si stava collegando un nodo distribuzione e lo si esegue in questo punto, non si visualizza alcuna modifica rispetto alla previsione originale. Sebbene il flusso ora unisce l'output del nugget del modello originale con i nuovi campi creati dalle analisi entità, lo stesso campo di previsione (*\$R-default*) nel modello dati non è stato aggiornato con le nuove informazioni.

Per fare ciò, si utilizzerà un nodo Riempimento che può sostituire i valori del campo. Collegare un nodo Riempimento al nodo Unione e aprire il nodo Riempimento.

Fare clic sul pulsante in alto alla destra **Riempi campi**, scorrere fino in fondo l'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, accertarsi che l'opzione **Sostituisci** è impostata su **In base alla condizione**, quindi nel campo **Condizione** immettere:

```
default != "default" and IsDuplicate = "Duplicate"
```

Nel campo **Sostituisci con**, inserire:

```
"default"
```

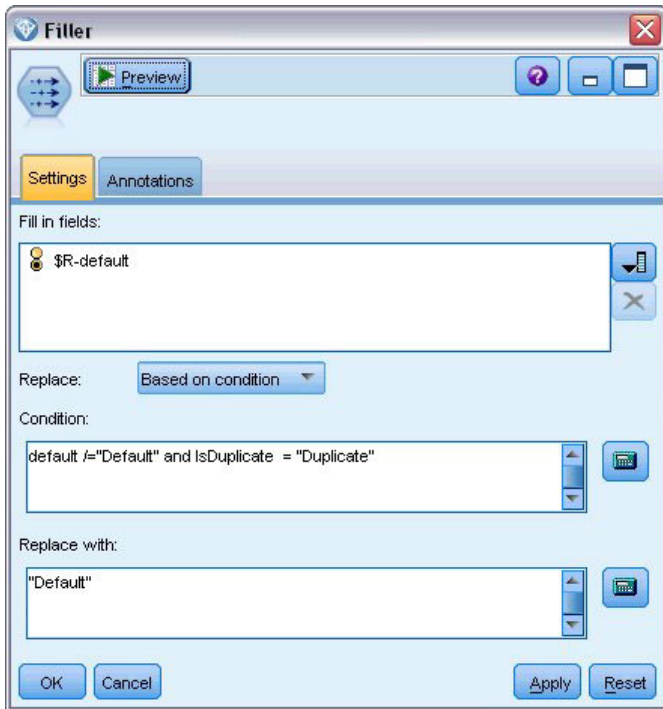


Figura 20. Indicazione della condizione per la sostituzione dei valori dei campi

Per queste impostazioni è necessaria una breve spiegazione. La condizione stabilisce che per ogni record in cui il valore del campo *default* del dataset originale non è uguale a **Default** e per il record che è stato contrassegnato come duplicato, il valore del campo *\$R-default* nel modello deve 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.

Si è finalmente pronti ad esaminare la differenza che le analisi entità hanno fatto. Dalla palette Grafici, collegare un nodo distribuzione al nodo Riempimento e aprire il nodo distribuzione. Fare clic sull'elenco **Campo** e scegliere **\$R-default**.

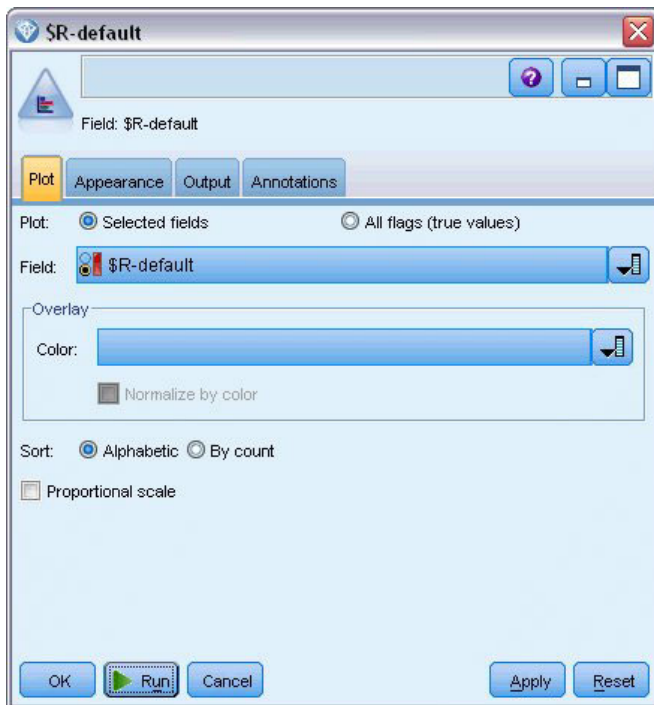


Figura 21. Impostazioni per il nodo distribuzione

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

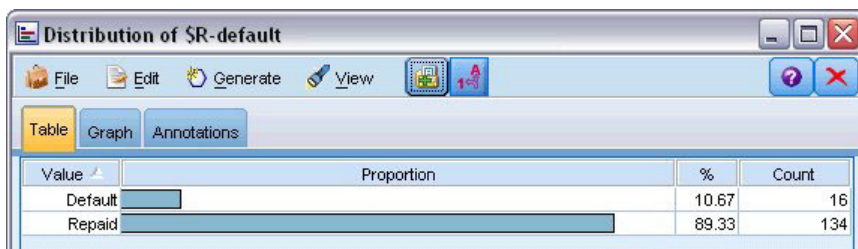


Figura 22. Output del nodo distribuzione dopo l'analisi delle entità

Ora esistono 16 richieste a rischio invece di 13. Queste richieste aggiuntive potrebbero essere molto costose se dovessero risultare inadempienti, quindi è possibile dimostrare graficamente alla banca il vantaggio di aggiungere analisi entità alle 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: Preferibilmente, si dovrebbero eliminare i record duplicati prima di eseguire qualsiasi altra elaborazione. Si potrebbe eseguire questa procedura utilizzando un nodo ADP (Automated Data Preparation) per analizzare i dati e individuare le correzioni, eliminare i campi problematici o non potenzialmente utili, derivare nuovi attributi quando necessario e migliorare le prestazioni attraverso tecniche di screening avanzate.

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

Appendice. Proprietà script per IBM SPSS Modeler Entity Analytics

Script con IBM SPSS Modeler Entity Analytics

Gli script di IBM SPSS Modeler Entity Analytics sono un potente strumento per automatizzare i processi nell'interfaccia utente. Tramite gli script è possibile eseguire gli stessi tipi di azioni eseguite con il mouse o la tastiera, nonché automatizzare le attività ripetitive o la cui esecuzione manuale richiederebbe un tempo molto maggiore. Per la descrizione relativa all'utilizzo degli script, consultare la guida *ScriptingAutomation.pdf* disponibile con IBM SPSS Modeler.

Proprietà comuni

Le proprietà che sono comuni ai nodi IBM SPSS Modeler Entity Analytics vengono elencate nella seguente tabella. Le informazioni relative a specifici nodi sono fornite nelle sezioni di seguito riportate:

Tabella 13. Proprietà comuni

Nome proprietà	Tipo di dati	Descrizione proprietà
entity_repository	['campo','campo', ... , 'campo']	La stringa di connessione al repository Formato: ['reposname', 'username', 'password'] Esempio: entity_repository = ['repos1', 'dba', 'psw1']
entity_type	stringa	Il tipo di entità (insieme di funzioni) da utilizzare. Esempio: entity_type = 'PERSON'

Proprietà entityanalytics_exportnode



Il nodo Esportazione EA è un nodo terminale che legge i dati dell'entità da un'origine dati e li esporta in un repository per la risoluzione dell'entità.

Tabella 14. Proprietà entityanalytics_exportnode

Proprietà entityanalytics_exportnode	Tipo di dati	Descrizione proprietà
mode	Add PurgeFirst	Modalità esportazione. Add aggiunge i record del file origine ai contenuti del repository esistenti; PurgeFirst rimuove i contenuti esistenti prima dell'esportazione.
source_tag	stringa	L'identificativo dell'origine dati. Esempio: source_tag = 'CUST'

Tabella 14. Proprietà entityanalytics_exportnode (Continua)

Proprietà entityanalytics_exportnode	Tipo di dati	Descrizione proprietà
unique_key_field	stringa	Campo di input da utilizzare per gli identificativi univoci per i record di dati. Esempio: unique_key_field = 'ID'
field_mapping	[['nome_campo' 'funzione.elemento' 'tipo_utilizzo']...]	Mappa i campi di input per la funzione corrispondente nel repository. Esempio: field_mapping = [['fname' 'NAME.GIVEN_NAME' ' '] ['addr1' 'ADDRESS.ADDR1' 'PRIMARY']] <i>Nota:</i> per impostare <i>tipo-utilizzo</i> all'equivalente di "(Automatico)", utilizzare ' ' come nel primo esempio sopra riportato.

Proprietà entityanalytics_sourcenode



Il nodo origine Entity Analytics(EA) legge le entità risolte dal repository e passa questi dati al flusso per l'ulteriore elaborazione, ad esempio la formattazione in un report.

Tabella 15. Proprietà entityanalytics_sourcenode

proprietà entityanalytics_sourcenode	Tipo di dati	Descrizione proprietà
source_tags	elenco	Elenco di tag per le origini dati che devono uscire dal repository. Esempio: source_tags=['LOANS', 'CUSTOMERS']
relationships	None Close All	Il criterio di corrispondenza per richiamare i dettagli delle relazioni dal repository. Nessuno non restituisce alcuna relazione. Chiudi restituisce corrispondenze strette a seconda dei dettagli come ad esempio il grado di separazione. Tutto restituisce tutte le relazioni possibili.
max_degree_separation	numero intero	Minimo 0, massimo 3.
output_entity_type	stringa	Elenco di tipi di entità utilizzate nel repository.

Proprietà entityanalytics_processnode



Il nodo Streaming di EA confronta i nuovi casi con i dati entità già presenti nel repository.

Tabella 16. Proprietà entityanalytics_processnode

proprietà entityanalytics_processnode	Tipo di dati	Descrizione proprietà
corrispondenza	Exact ByIdentifier All	Il criterio di corrispondenza per richiamare le entità dal repository. Exact restituisce solo le corrispondenze esatte. ByIdentifier restituisce le corrispondenze esatte e le entità che condividono gli stessi identificativi. All restituisce tutte le possibili corrispondenze
save_search_records	booleano	
relationships	None Close All	Il criterio di corrispondenza per richiamare i dettagli delle relazioni dal repository. Nessuno non restituisce alcuna relazione. Chiodi restituisce corrispondenze strette a seconda dei dettagli come ad esempio il grado di separazione. Tutto restituisce tutte le relazioni possibili.
max_degree_separation	numero intero	Minimo 0, massimo 3.
output_entity_type	stringa	Elenco di tipi di entità utilizzate nel repository.

Note

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

È possibile che IBM non offra i prodotti, servizi o funzioni illustrati in questa documentazione. Consultare il rappresentante locale IBM per le informazioni sui prodotti e servizi attualmente disponibili nella propria zona. Qualsiasi riferimento a un prodotto, programma o servizio IBM non implica o intende dichiarare che può essere utilizzato solo quel prodotto, programma o servizio IBM. In sostituzione a quelli forniti da IBM è possibile utilizzare qualsiasi prodotto, programma o servizio funzionalmente equivalente che non comporti la violazione dei diritti di proprietà intellettuale IBM o altri diritti. Tuttavia, è responsabilità dell'utente valutare e verificare il funzionamento di qualsiasi prodotto, programma o servizio non IBM.

IBM potrebbe avere brevetti o domande di brevetti in corso relativi ad argomenti discussi nella presente pubblicazione. 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 Europe
Schoenaicher Str.220
D-7030 Boeblingen
Deutschland

Per richieste di licenze relative ad informazioni double-byte (DBCS) contattare il Dipartimento di Proprietà Intellettuale IBM nel proprio paese o inviare richieste per iscritto a:

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

Il seguente paragrafo non è valido nel Regno Unito o per tutti i paesi le cui leggi nazionali siano in contrasto con le disposizioni in esso contenute INTERNATIONAL BUSINESS MACHINES CORPORATION FORNISCE QUESTA PUBBLICAZIONE "NELLO STATO IN CUI ESSA SI TROVA" SENZA ALCUNA GARANZIA ESPLICITA O IMPLICITA IVI INCLUSE EVENTUALI GARANZIE DI COMMERCIALIZZABILITÀ ED IDONEITÀ AD UNO SCOPO PARTICOLARE 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 si riserva il diritto di apportare miglioramenti e/o modifiche al prodotto o programma descritto in questa pubblicazione in qualsiasi momento e 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 disponibili sui siti Web non fanno parte di questo prodotto IBM e l'utilizzo di questi è a discrezione dell'utente.

IBM può utilizzare o distribuire qualsiasi informazione fornita dall'utente nel modo che ritiene più idoneo senza incorrere in alcun obbligo nei confronti dell'utente stesso.

I licenziatari di questo programma che desiderano avere informazioni allo scopo di abilitare: (i) lo scambio di informazioni tra i programmi creati indipendentemente e gli altri programmi (incluso il presente) e (ii) il reciproco utilizzo di informazioni che sono state scambiate, dovrebbero contattare:

IBM Software Group
ATTN: Licensing
200 W. Madison St.
Chicago, IL; 60606
U.S.A.

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

Il programma su licenza descritto in questa documentazione e tutto il materiale su licenza ad esso relativo vengono forniti da IBM nei termini del Customer Agreement IBM IBM International Program License Agreement o di eventuali accordi equivalenti intercorsi 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 testato quei prodotti e non può garantire 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 di business 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 e ibm.com sono marchi o marchi registrati di International Business Machines Corp., registrati in molte giurisdizioni nel mondo. Altri nomi di prodotti e servizi possono essere marchi di IBM o altre società. Un elenco aggiornato dei marchi IBM è disponibile sul Web nella sezione "Copyright and trademark information" all'indirizzo .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 logo basati su Java sono marchi o marchi registrati di Oracle e/o suoi affiliati.

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

Indice analitico

A

- analisi entità
 - confronto con analisi predittiva 2
 - definizione 1
 - utilizzo con gli altri prodotti IBM SPSS 31
 - utilizzo con IBM SPSS Modeler 5
- anonimizzazione delle funzioni repository entità 22
- assegnazioni delle porte
 - configurazione per l'analisi delle entità 32

C

- caratteristiche
 - repository entità 7, 16, 17, 18, 19, 20, 21, 29, 30
- chiavi univoche
 - analisi entità 7
 - repository entità 17
- configurazione
 - repository entità 18, 25, 26
- corrispondenza entità, impostazione soglia 25
- credenziali amministratore
 - gestione per analisi entità 33

E

- eliminazione
 - repository entità 36
- eliminazione di dati non utilizzati
 - repository entità 35
- esportazione
 - dati a un repository di entità 7

I

- identità, risoluzione con Entity Analytics 7
- identità risolte, analisi con le analisi di entità 26
- informazioni sui tipi, impostazione per l'analisi delle entità 27

M

- mappatura di campi
 - a funzioni del repository entità 7, 16, 17, 18, 19, 29, 30

N

- nodi
 - aggiunta a un flusso di analisi delle entità 28

- nodi di elaborazione
 - analisi entità 28
- nodi di esportazione
 - analisi entità 7, 13
- nodi di processo
 - analisi entità 10
- nodi origine
 - analisi entità 9, 26
- Nodo Esportazione EA, analisi entità 13
- Nodo origine Entity Analytics(EA) 26
- Nodo Streaming di EA, analisi entità 28
- nuovi casi, confronto rispetto al repository analisi entità 28

O

- origine dati
 - connessione con analisi entità 6, 14
 - visualizzazione per analisi entità 30
- output
 - dall'analisi delle entità 31

P

- proprietà
 - script 55
- proprietà flusso
 - impostazione per le analisi entità 34

R

- regole di risoluzione, analisi entità 25
- repository
 - amministrazione dell'analisi delle entità 32
 - analisi entità 6, 7, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 25, 26, 28, 29, 30, 35, 36
 - directory di archiviazione per l'analisi entità, modifica 33
- repository entità 13
 - anonimizzazione 22
 - attività amministrative 32
 - caratteristiche 21
 - configurazione 18, 25, 26
 - configurazione delle assegnazioni delle porte 32
 - confronto nuovi casi 28
 - connessione a IBM SPSS Modeler 7
 - creazione 6, 14, 15
 - eliminazione 36
 - eliminazione di dati non utilizzati 35
 - gestione 20
 - gestione credenziali amministratore 33
 - impostazione 13
 - impostazioni proprietà flusso 34
 - opzioni 16
 - spostamento ad una diversa directory di archiviazione 33

- repository entità (*Continua*)
 - svuotamento 35
- ridenominazione
 - campi dati per analisi entità 27
- risoluzione di identità, Entity Analytics 7

S

- script
 - proprietà 55
- soglia per la corrispondenza delle entità, analisi entità 25
- sorgente dati, selezione per analisi entità 26
- sorgenti dati
 - visualizzazione per analisi entità 16
- svuotamento
 - repository entità 35

T

- tag sorgente
 - repository entità 17
- tipi di entità
 - analisi entità 23
 - repository entità 17
- tipi di utilizzo, analisi entità 23



Stampato in Italia