

IBM SPSS Modeler

透過視覺化的資料科學以及機器學習，
加速創造價值的進程



特色

- 發揮視覺化及程式化資料科學的力量 -- 與 IBM® SPSS® Modeler for IBM Watson™ Studio 攜手合作，或是從訂閱部署 IBM SPSS Modeler 開始
- 在多雲環境中包含的文字、純文字檔、資料庫、資料倉儲及 Hadoop 分佈等更多資料來源中，運用這些資料發揮價值並找出模式。
- 將 40 種以上現成可用的機器學習演算法運用於工作中，讓模型開發及管理更輕鬆
- 與 Apache Spark 整合以進行快速記憶體內運算
- 透過程式化、視覺化以及其他技巧，優化資料科學及業務團隊的生產力
- 透過資料庫內效能及最少的資料移動，加速資料分析
- 透過開放程式碼技術，例如 R、Python、Spark 以及 Hadoop，就能擴展您的能力 -- 不一定非得撰寫程式碼

概觀

在現今高度競爭的商業環境中，各行各業領導者正在微調對於資料科學的投資，以改善商業成果。這些成果包括：

- 獲得及留住客戶
- 促進營運效率
- 將詐騙降至最低並減輕風險
- 優化招聘流程並減少人員流失
- 建立新的商業模型

資料科學是使用分析技術發掘資料中的模式及應用其結果來獲得商業價值的程序。使用描述分析、預測建模、文字分析、地理空間分析、實體分析、決策管理以及優化功能，來識別模式並將預測模型部署到作業系統中。

系統與人員可以使用這些模式及模型來找出洞察，而能持續在影響點中做出正確決策。成果的優化程度取決於預測情報，隱藏於規模及複雜度不斷提高的資料中。

IBM SPSS Modeler 是強大的視覺化資料科學以及機器學習工具，主要設計來為個人、團隊小組、系統以及您的企業在制定決策時帶入預測情報。從小範圍的桌面部署到大範圍的作業系統整合，SPSS Modeler 都能提供您一系列先進的演算法以及技術。將這些技術應用於決策，能帶來 ROI 的快速成長，並讓組織能主動並不斷地降低成本且提高生產力。



資料表

IBM SPSS Modeler 可透過訂閱或永久的授權版本來滿足將近百分百的企業需求：

IBM SPSS Modeler 提供作為 **IBM Watson Studio** 中的一部分，為一互動式的協同平台，其中資料科學家能夠使用多個開放程式碼以及 **IBM** 工具來激發他們的洞察。您的程式語言筆記本以及視覺化資料串可以在單一視圖中顯示，以便能輕鬆地選擇正確模型進行訓練並加以部署。

IBM SPSS Modeler 訂閱提供您所需的功能，讓您能存取並匯出結構化資料集而不受資料大小限制、使用 40 種以上的演算法分析資料，以及使用您的桌面部署結果。提供彈性的逐月付款選項，而不需承諾訂購一整年，此外您可以選擇加入 SQL 優化或是文字分析功能。

IBM SPSS Modeler Professional 軟體包括利用結構化資料所需的工具，這些資料包括您的 CRM 系統中所追蹤記錄的行為與互動、個人背景資訊、採購行為以及銷售資料等 -- 在多使用者的環境中。

IBM SPSS Modeler Premium 軟體除了結構化資料分析，還包括非結構化資料，例如網路活動、部落格內容、客戶意見回饋、電子郵件、文章等，以便能建立更準確的預測模型。

IBM SPSS Modeler Gold 軟體包含 IBM SPSS Modeler Premium 軟體的所有功能，還加上建置預測模型並直接部署到商業流程的功能。將預測分析與組織流程和作業系統內部的規則、評分以及優化相結合，能在影響點送出建議動作。

不論您選擇哪一種 SPSS Modeler 解決方案，您都能夠：

- 存取所有類型的資料
- 拓展分析範圍
- 透過靈活的部署符合所需

存取所有類型資料

資料有眾多來源且呈指數比率大量發生，因而可為組織注入新資訊和前所未有的機會，讓組織得以駕馭及利用資料並實現其價值。此類資料以不同程式儲存於各種系統中，因此整合可謂一大挑戰。資料量如此龐大，您無法對其進行手動分析，也無法從報告中的表格查出事件是否會發生的原因。分析流程還有另一個挑戰，癥結點在缺乏利用資料來擷取其價值的專業分析師。

透過 **SPSS Modeler**，您與您的組織能夠使用可用的資料（包括結構性、非結構性或是空間性的資料），並透過發掘未觸及的機會及新興資訊擷取資料的價值。從資料中獲得的新洞察，讓您得以預測可能發生的情況、轉為主動方並且優化成果，不再僅僅是被動地做出反應。

SPSS Modeler 讓您能夠使用一系列的分析技術存取資料來源，例如文字、純文字檔、資料庫、資料倉儲以及 **Hadoop** 分佈。這些統計資料技術使用歷程資料做出現況或未來事件的預測。另外包括資料存取、資料準備、資料建模以及互動式視覺化的功能。經由準備以及建模的自動化程序，可適用於廣泛的分析能力。

SPSS Modeler 直覺式的圖形介面讓使用者得以將資料科學程序中的每個步驟視覺化，作為串流的一部分。透過與這些串流互動，分析師與商業使用者能夠分工合作，讓商業知識以及領域專門知識加入資料科學程序中。使用者能夠專注於發掘洞察，而不是撰寫程式碼這類技術作業上。他們還能夠專注思緒做分析，並且更深度地探索資料，而這兩者都能為您的組織發掘出更多有意義的關係。

資料準備及操作

分析資料前的準備工作相當重要，但是在分析工作中，是費時的步驟。SPSS Modeler 加速資料準備讓此處理程序變得輕鬆容易，幫助您確保資料處在適於分析的最佳格式。此自動化作業包括分析資料以及識別修正程式、篩選欄位、適時衍生新屬性，並且透過智慧型篩選技術協助改善效能。

SPSS Modeler 提供各種方式，在記錄或欄位（或變數）層次操作並準備資料以進行分析。為了幫助確認您的資料為所需分析之特定類型的最佳格式，SPSS Modeler 使用以下方法：

- **記錄作業。**「選取 (Select)」、「抽樣 (Sample)」及「區別 (Distinct)」節點讓您能夠選擇特定的資料列。您可以合併及附加節點，藉由新增直欄或列到資料集來結合資料。「彙總」以及「最近消費時間、消費次數、消費總金額 (RFM)」節點會把記錄彙總到單一系列。
 - 「平衡」節點可調整不平衡資料中的記錄比例，「排序」節點則根據值重新排序資料。「空間時間方框 (STB)」節點會建立記錄的地理空間以及時間型資料。
- **欄位作業。**「類型」節點能指定資料集的 meta 資料及內容，而「過濾」節點能夠捨棄欄位。「衍生」節點能建立新的欄位，「填充值」節點則能夠取代現有欄位值。可以由「設定通知」、「重組」或「轉置」節點重組資料，並使用「重新分類」或「收納」節點重新分組。如要協助建模，「分割」節點可以分割資料，「歷程」節點以及「時間間隔」節點則可建立額外欄位。「欄位重新排序」節點能定義顯示的排序，便於檢視特定欄位。

透過一系列的技術拓展您的分析範圍

分析技術持續演進，為分析師提供多種選擇，在問題浮現之前就能著手處理。此外，由於不斷發展的技術以及成為可用的新類型資料（例如手機的位置資料或是 IoT 裝置的即時資料），如何運用此資料的各種相關問題與挑戰也應運而生。因此，必須要有創新的技術。

使用 SPSS Modeler，您的分析師可以使用解決方案來解決他們的商業問題，此解決方案主要設計來處理簡單的描述分析、最複雜的優化問題，以及介於這兩種程度之間的所有問題。SPSS Modeler 的特色是其功能超越現今分析師以及資料科學家的標準分析需求。有 30 種以上的一系列模型、視覺化拖放建模、內建資料準備、文字分析、實體分析、社交網路分析，以及在平行處理程序上建置模型的能力，都能協助您處理最複雜的問題。

資料科學視覺化拖放建模

透過 SPSS Modeler 的拖放建模功能，即使不是分析師，無須具備特殊技能也能夠快速產出準確的模型。此外，進階的預測建模功能讓專業的分析師及資料科學家能夠建立最複雜的串流資料。

視覺化建模讓您得以比較多個建模方式。透過對每個模型類型設定特定選項（或是使用預設值），您可以探索大量的模型組合及選項。產出的模型接著會按照指定的方式排序，留存最佳方式以供評分或供未來分析使用。

一系列的模型

SPSS Modeler 提供一系列建模技術，包括下列演算法：

- **分類演算法**。根據歷程資料搭配科技，做出預測或預報。範例包括決策樹、隨機樹、神經網路、邏輯迴歸、支援向量機、Cox 迴歸、一般性線性混合模型等等。針對二進位及數值成果使用自動分類建模，以簡化模型建立或自學回應建模，來建立可持續更新或不必重建模型就可以重新評估的模型。
- **分割演算法**。使用技術將人員分組或是偵測異常的模式，此類技術如自動叢集、異常偵測以及叢集神經網路。使用自動分類以單一步驟來套用多個演算法，不再需要費時猜測選擇適用的技術。
- **關聯分析演算法**。使用 Apriori、CARMA 以及序列關聯探索關聯、鏈結或序列。
- **時間序列與預測**。藉由統計建模技術，針對一或多個序列在經歷一段時間後產出預報。使用時間因果建模，您可以在大量序列中探索因果關係。
- **R 與 Python 程式設計語言的可擴充性**。使用 R 應用轉換、使用 Script 分析、彙總或產出文字及圖形輸出。透過「自訂對話建置器」，您可與選擇不使用程式設計進行分析的人員，共用並重複使用 R 及 Python 和 Python 程式碼。
- **蒙地卡羅模擬分析**。解釋輸入資料到預測模型中的不確定性。模型不確定的輸入資料是基於歷程資料或透過機率分佈來產生模擬值，接著將其使用於預測模型來產生成果。結果是成果的分佈，能針對問題根據現實產生的資料提出答案。

- **實體分析**。識別關係並透過解決記錄本身的身分衝突改善現行資料的連貫性與一致性。在許多領域中，識別這些關係至關重要，包括客戶關係管理、詐騙偵測、反洗錢以及安全性。
- **社交網路分析**。探索社交實體之間的關係，以及這些關係對個人行為的潛在影響。這些功能能夠幫助將關係的相關資訊轉換為關鍵績效指標，顯示個人與群體的社會行為。

地理空間分析

使用 SPSS Modeler 可讓您探索與位置關聯的資料元素關係，並對您所擁有的資料進行地理空間分析，以獲取在圖表或表格中不可見的洞察。透過空間分析，您可以輕鬆地使用 ESRI Shape 檔來併入地理空間資料。藉由分析非空間及空間資料，可提升整體模型的準確性，而您將能夠對於人與事件獲得更深入的洞察。

透過探索空間及非空間屬性間的關聯規則，在您的分析中增加新的維度。使用時空預測，您可找出 2D 空間位置測得的適用線性模型，讓您能夠輕鬆預測熱門區域以及那些區域在不同時間的變化情形。您可將此技術應用於發掘地理空間資料的相關領域，例如犯罪模式分析、傳染病監測、建築管理以及分支機構效能分析。

文字分析

SPSS Modeler 互動式的視覺化環境，使用進階的語言學技術以及「自然語言處理程序」，快速處理非結構化文字資料。從這些文字擷取並組織主要概念。可自訂的行業特定文字分析套件讓您能夠分析相關術語及詞組，以及在正確情境中的字首語、表情符號及俚語。互動式圖形協助您探索並顯示文字資料及模式，進行即時分析。您可建立階層式分類結構，並將其納入作為預測模型的輸入資料，以獲取更好更準確的決策與結果。預先定義的種類，例如階層關係、註釋以及關鍵字描述子，可匯入來對起始非結構化資料進行分類，讓您能夠根據更多的細節，以更具邏輯性的方式組織概念。

透過靈活的模型部署符合所需

在您的組織內進行分析部署取決於許多環境因素。這些因素包括必須處理的商業問題、您選擇的作業系統及平台，以及基礎架構中的其他技術及資料來源。技術部分，尤其是軟體，必須具備充分的靈活度以滿足各種排列而仍能提供預期的效能與結果。

SPSS Modeler 架構為開放式，支援各種平台及語言，包括 R 及 Python。可提供作為 Watson Studio 的一部分，也能獨立使用或作為訂閱型軟體產品使用。

您可以在本機、雲端部署 SPSS Modeler，或甚至使用混合式雲端方式；或者，也可以接著滿懷信心地搭配現有系統用來協助優化效能並處理商業問題。此靈活部署能藉由排程或隨需應變提供結果予人員與流程，以消弭分析與實作間的差異。SPSS Modeler 串流能夠根據模型更新的用途進行部署，或是自動工作排程。開發人員可使用內嵌的預測建模 Java API 將 SPSS Modeler 串流內嵌到 Java™ 應用程式。而且您可以使用決策管理或其他預測應用程式來進行部署。

決策管理

決策管理將 SPSS Modeler 的預測功能延伸到日常的商業流程，作為前線員工及系統的後盾。它將預測模型、簡單的規則以及評分整合到您的系統，以自動化、管理及優化大量的決策。

接著在人員需求之所在與時機建議動作，例如在客戶電話會談中進行交叉銷售、決定索賠的最佳方式、使用公用事業來分配頻寬或對電子商務客戶提供優惠。數千個決策可以在作業層次完成，且與您的組織之目標與策略完全一致。

藉由 SPSS Modeler 的決策管理功能，您可以：

- 預測模型能預見最有可能的成果並識別導致該成果的因素，例如客戶對於推出之優惠的回應傾向或提出的索賠為詐騙的風險。
- 商業規則能將諸如商業原則或法律及合規性等元素所決定的參數進行自動化。SPSS Modeler 直接提供基本規則支援。為能取得符合企業層面需求相應更加健全的規則，也支援與 IBM Decision Optimization 的整合。
- 整合評分能對正確的人員及系統提供近乎即時的建議，因此可為所有事業單位做出顧全資源並符合策略的決策。

優化

即使是最盡心盡力的組織，也無法為每位顧客省去面對前線作業的等待時間，保險業者也無法針對每件索賠進行調查。商業運作受到現實世界的限制，取決在充足的人力、設備以及投資限制。優化讓組織能夠藉由識別符合特定目標的最佳解決方案，善用珍貴的資源。例如，將市場行銷活動的收益提升至最高，或將詐騙或顧客流失風險降至最低。優化可隨需應變，例如針對個人提出最佳優惠，或是對所有符合資格的客戶分批提供優惠。

執行及排程

SPSS Modeler 包括設計來使用自動化的功能，為您帶來更加一致的結果。更完善的一致性能加強人員對於分析的信心，因為管理層能夠有效地控管分析流程發生所在的商業環境。這樣的控管協助確保滿足所有內部及外部程序化需求。

藉由 SPSS Modeler，您的分析師能夠構建靈活、可重複的分析流程使其更具操作性，也就是說，起始於正確時間點並且與其他企業流程整合。預先定義的模型管理處理程序能夠幫助模型保持相關性及準確度。

資料庫內功能

SPSS Modeler 提供許多功能，可用最少的資料移動，將分析推送到資料庫內，例如：

- **SQL 回推 (Pushback)**。透過 SPSS Modeler Server，即使是在 IBM Z® 環境中，不需要從大型資料庫移動資料，因為分析可以在資料庫中進行。「SQL 回推」在資料庫內可進行資料轉換及準備，而不需撰寫任何 SQL 或進行其他的程式設計。使得分析效能大幅提升。
- **資料庫內評分**。特定的資料庫評分配接器，在 IBM SPSS Modeler with Scoring Adapter for zEnterprise®、IBM Db2®、IBM PureData® System for Analytics (搭載：IBM Netezza®) 以及 Teradata 解決方案中可取得，擴充 SPSS Modeler 演算法的範圍，這些演算法可在資料庫內評分，進一步減少評分之前擷取資料的需求。
- **資料庫函數**。SPSS Modeler 可使用絕大部分的使用者定義函數、資料庫彙總及資料庫提供的時間範圍型彙總函數。這些都由 SPSS Modeler 工作台公開以延伸可用的原生功能並確保「SQL 回推」。

- **資料庫內分析**。SPSS Modeler Server 支援整合分析功能、建模工具以及資料庫原生演算法，可在 IBM Db2 Analytics Accelerator on Z Hybrid Transactional and Analytic Processing、PureData System for Analytics、Oracle Data Miner、Microsoft® Analysis Services 等中取得。您可以在資料庫內建置、評分並儲存模型 -- 只需要在 SPSS Modeler 工作台進行。

與 IBM 技術整合

除了提供作為 Watson Studio 的一部分，SPSS Modeler 包含匯出資料到 IBM Cognos® Business Intelligence 及 Cognos TM1® 軟體的功能。分析結果可針對僅需結果的主要決策者分配為報告、監控及規劃部分。當需要進一步分析時，SPSS Modeler 也能夠將其作為資料來源存取，表示流程可再次繼續進行，從而將結果回饋到問題開始所在。

IBM SPSS Statistics 提供執行進一步統計資料分析及資料管理的能力，藉由節點的專用區段讓 SPSS Modeler 及其資料科學能力更加完善。

SPSS Modeler 提供 PureData System for Analytics 的支援，以便從 SPSS Modeler Interface 存取特定的模型並善用硬體的速度與效能。

針對大數據分析，SPSS Analytic Server 作為 SPSS Modeler 增補，可在 Hadoop 分佈中處理分析。

SPSS Modeler 與 Apache Spark 透過 SPSS Analytic Server 的整合提供大量大數據演算法以及記憶體內處理功能，便於建立快速及有效的機器學習模型。

SPSS Modeler 串流也可以搭配 IBM InfoSphere® Streams 部署，以用於需要預測評分的高速串流資料應用程式。

選擇 IBM SPSS Modeler 的理由

SPSS Modeler 作為商業後盾加速創造價值的時間進程、提高資料科學與商業團隊的生產力，並確保在整個資料科學生命週期中控管與安全需求相符。可將 SPSS Modeler 提供作為 Watson Studio 的一部分，用以開始實行資料科學 -- 包括測試、開發、部署等。您的組織可使用 SPSS Modeler 執行分析而不需考慮資料所在位置（內部部署或在雲端）、資料大小，或是否為結構性資料。可調式主從架構讓使用者得以存取所有資料，不論這些資料是純文字檔格式或是在大數據環境中。分析結果會被回推到原始檔執行，將資料移動降到最低限度以提高效能。

藉由 SPSS Modeler，各種專業領域的使用者得以分工合作共同面對目標使用案例，快速地產出更好的成果。所提供的分析技術涵蓋範圍從預測規範分析到進階演算法，包括自動化建模、文字分析、實體分析、社交網路分析、決策管理以及優化。直覺式的介面設計為服務廣泛的使用者類型，從非技術型商業使用者到分析專家。SPSS Modeler 短暫的學習曲線對於入門及進階使用者來說很有吸引力，因此他們能迅速發掘洞察並實現商業結果。

進一步瞭解

進一步瞭解 IBM SPSS Modeler 或到下列網址開始免費試用：

ibm.com/products/spss-modeler

善用我們活躍、持續成長的開放原始碼社群，您可以在其中找到協助您拓展應用 IBM 預測分析軟體的資源。資源包括部落格、影片、指導教學，以及廣泛的程式庫，包含 6,000 種以上的預測擴充功能，協助您運用熱門的程式設計語言，例如 R、Python 及 Java。社群網址為：

developer.ibm.com/predictiveanalytics/



© Copyright IBM Corporation 2018

IBM Corporation
1 New Orchard Road Armonk,
NY 10504

2018 年 8 月美國生產

IBM、IBM 標誌、ibm.com、Cognos、Db2、IBM Z、InfoSphere、PureData、SPSS、TM1、Watson 和 zEnterprise 為 International Business Machines Corporation (IBM) 在全球許多司法管轄區註冊的商標。其他產品與服務名稱可能為 IBM 或其他公司的商標。IBM 商標最新清單可於下列網站之「著作權與商標資訊」(Copyright and trademark information) 網頁上取得：ibm.com/legal/copytrade.shtml

Netezza 是 IBM 旗下公司 IBM International Group B.V. 的註冊商標。

Microsoft 是 Microsoft Corporation 在美國及/或其他國家或地區的商標。

Java 和所有以 Java 為基礎的商標和標誌是 Oracle 及/或其關係企業的商標或註冊商標。

此文件現行日期為發佈起始日期，IBM 可能隨時變更之。並非所有產品與服務都會在 IBM 營運的每個國家或地區供應。

其他任何產品或程式與 IBM 產品和程式一起使用時，使用者需自行負責評估與驗證。本文中的資訊僅「依現狀」提供，不含任何明示或默示之保證，包括不含適售性、符合特定效用及未涉侵權。IBM 產品之保證，係依據以提供該等產品之合約條款提供。



請回收