



## 主要特点

- 访问平面文件、数据库、数据仓库和分布式 Hadoop 等多种数据源，发掘数据隐含模式
- 实现专门用于高效处理大数据并为大数据建模的机器学习算法，与 Apache Spark 集成，用于在内存中快速进行计算
- 在关键时刻即时向人员和系统提供符合资源现状和公司策略的预测型决策
- 单一平台专门用于处理从简单的描述性分析到最复杂的优化问题的一切事务，协助用户解决商业问题
- 在数据库中进行分析，最大限度降低数据传送量，在充分利用原有 IT 投资的同时，用更短时间分析海量数据
- 充分利用一款开放式平台，该平台可在大多数环境中部署，并且可集成其他 IBM 解决方案，以填补分析与行动之间的空档

# IBM SPSS Modeler

提供智能预测，优化决策制定，  
实现更佳业务成果

## 概览

在商业环境中，分析的主要目的是提高业务成果。这些成果包括：

- 降低客户流失率，增加收益
- 利用呼叫中心，提高交叉销售率
- 提前识破欺骗性索赔，降低成本
- 在生产线上进行部件检修，最大限度缩短宕机时间

采用分析技术优化决策，所得成果往往更佳。

数据挖掘流程利用分析技术揭示数据中蕴含的模式规律。描述性分析、预测性建模、文本分析、地理空间分析、实体分析、决策管理与优化技术常用于识别数据模式以及在操作系统中部署预测性模型。系统和相关人员可利用这些模式和模型获得分析数据，在关键时刻始终如一地作出合理决策。利用智能预测技术对日益庞大且复杂化的数据进行分析，可最大限度改进业务成果。

IBM® SPSS® Modeler 是一款功能强大的预测性分析平台，专门用于提供智能预测，优化个人、团队、系统和企业的决策。SPSS Modeler 具备从桌面部署一直到与操作系统集成的扩展性，为用户提供一系列高级算法和技术。运用这些技术优化决策，不但可缩短投资回报周期，而且有助于企业主动缩减成本，同时不断提高生产力。



SPSS Modeler 有四个版本，几乎可满足任何分析需求：

### IBM SPSS Modeler Professional

揭示结构化数据中的隐含模式。SPSS Modeler Professional 可提供高级算法、数据处理和自动化建模与准备技术，建立有助于改进业务成果的预测性模型。

### IBM SPSS Modeler Premium

利用一系列高级算法和功能（如文本分析、实体分析和社交网络分析）以及自动化建模与准备技术，从几乎任意类型的数据中获取分析数据。

### IBM SPSS Modeler Gold

利用决策管理功能直接在业务流程中建立与部署预测性模型，使相关人员与系统每次都可以作出合理的决策。利用 SPSS Modeler Gold on Cloud，可在网上按月订阅这些功能。

### IBM SPSS Predictive Analytics Enterprise

利用一款包含 SPSS Modeler 的多元预测性分析解决方案，在关键时刻优化决策制定。

每个版本的 SPSS Modeler 均支持：

- 访问各类数据。
- 拓宽分析范围。
- 灵活部署，满足个性化需求。

## 访问各类数据

大量数据源的数据生成速率已经达到指数级，从而为有能力利用这些数据并实现其价值的企业提供全新信息和潜在机遇。这些数据以各种格式存储在不同系统中，因此可能很难进行汇总。另外，面对如此庞大的数据量，根本无法进行人工分析，也无法通过查阅报表中的表格，找到某事件发生与否的根本原因。分析流程还面临另一项挑战，即缺少可从数据中提取出商业价值的经验丰富的分析师。

借助 SPSS Modeler，个人用户和企业用户可利用已有数据（空间和非空间数据），通过挖掘潜在机遇和全新信息，从这些数据中提取商业价值。利用从数据中获得的全新分析数据，可以预测未来趋势、主动出击并优化成果，而不是仅仅根据现有状况作出反应。

SPSS Modeler 支持使用一系列分析技术访问数据源，如平面文件、数据库、数据仓库和分布式 Hadoop。这些统计技术利用历史数据对现有情况或未来事件作出预测。SPSS Modeler 还支持数据访问、数据准备、数据建模和交互式可视化功能。采用自动化的数据准备和建模程序，适用于多种分析用途。

SPSS Modeler 拥有直观的图形界面，用户可直接查看数据挖掘流程的各个步骤，每个步骤可作为“串流”的一部分。通过与这些“串流”进行交互，分析师和商业用户可紧密合作，为数据挖掘流程补充商业知识和行业经验。用户可专注于洞察发掘，而非诸如代码编写之类的技术性工作。用户也可进行“思路”分析和深层次数据探究，揭示对企业发展有重大意义的其他关系。

## 数据准备和操作

准备分析数据是分析流程中很重要也很耗时的一步。SPSS Modeler 支持自动化数据准备，可简化数据准备流程并帮助确保数据处于最佳分析格式。自动化任务包括分析数据和识别修正、筛选字段、获取新属性（必要时）以及通过智能筛选技术优化性能。

SPSS Modeler 支持以多种方式在记录或字段（或变量）级别操作和准备分析数据。对于所采用的特定类型的分析而言，确保数据处于最佳格式的方法包括：

- **记录操作。** Select、Sample 和 Distinct 节点支持选择特定行的数据。通过将列或行添加到数据集中，可以合并和添加节点，以用于数据拼接。Aggregate 和 Recency、Frequency、Monetary (RFM) Aggregate 节点可将记录汇总到单行中。Balance 节点可调节非平衡数据中的记录比例，Sort 节点可根据值进行重新排序。Space Time Box 节点可创建基于地理空间和时间的记录数据。
- **字段操作。** Type 节点可指定数据集的元数据和属性，Filter 节点可删除字段。Derive 节点可创建新字段，Filler 节点可替换现有字段值。通过 Set to Flag、Restructure 或 Transpose 节点可调整数据结构，通过 Reclassify 或 Binning 节点可对数据重新分组。Partition 节点可拆分数据，History 节点和 Time Intervals 节点可创建附加字段，为建模提供帮助。Field Reorder 节点可定义显示顺序，使特定字段更易于查看。

## 利用一系列技术拓宽分析范围

随着分析技术的不断革新，分析师们有数不胜数的方案来处理所面临的问题。另外，随着技术的进步和新型数据的出现（如手机或信号塔的位置数据），对于如何以最合理的方式利用该数据，随之产生出多种问题和挑战。因此，亟需一种创新型的技术。

借助 SPSS Modeler，分析师可利用单一平台解决当前的商业问题，该平台专用于解决从简单的描述性分析到最复杂的优化问题的一切问题。SPSS Modeler 具有超越当今分析师标准分析要求的功能。借助一系列模型、自动化建模与数据准备、文本分析、实体分析、社交网络分析以及并行建模功能，用户最棘手的问题都将迎刃而解。

### 自动化数据建模

利用 SPSS Modeler 的自动化建模功能，快速创建准确的数据模型不再是分析师的专利，不具备任何特殊技能的普通人也能轻松做到。另外，专业分析师可利用高级预测建模功能创建最复杂的串流。

自动化建模支持将多种建模方案进行比较。通过设置各种模型的特定选项（或使用默认值），用户可尝试使用多种模型组合与选项。然后根据特定测量值评价所生成的模型，保存最佳模型，供评分或进一步分析使用。

## 一系列模型

SPSS Modeler 提供多种建模技术，包括以下算法：

- **分类算法。**利用多种技术，根据历史数据作出预测或预报。技术示例包括决策树、随机树、神经网络、逻辑回归、支持向量机、Cox 回归分析和广义线性混合效应模型 (GLMM) 等。使用自动分类建模获得二进制和数值结果，以简化模型创建，或使用自主学习反应建模 (SLRM) 建立可持续更新的模型，或无需重建即可重新评估的模型。
- **分割算法。**利用自动群集、异常检测和群集神经网络等技术进行人员分组或异常模式检测。利用自动分类功能在单个步骤中应用多种算法，并在选择合适技术的过程中摒弃臆测成分。
- **关联算法。**利用 Apriori、CARMA 和序列关联算法发掘关联、连接或序列关系。
- **时间序列与预测。**采用统计建模技术针对一个或多个时间序列生成预测。利用时序因果建模法可揭示众多序列之间的因果关系。
- **R 和 Python 编程语言的可扩展性。**利用变换算法和/或脚本，通过 R 语言分析、总结或产生文本和图形输出。通过 Custom Dialog Builder，可以与不采用编程分析法的人员共享和复用 R 和 Python 语言以及 Python 代码。
- **Monte Carlo 模拟。**说明预测性模型输入的不确定性。模型不确定性输入根据历史数据或概率分布生成模拟值，然后将模拟值用于预测性模型，进而得出结果。该结果是结果分布图，可以解答根据实际数据产生的问题。

- **实体分析。**通过解决数据记录本身存在的身份冲突，识别现有数据中存在的关系，并提高现有数据的连贯性和一致性。识别这些数据关系在许多领域都极其重要，其中包括客户关系管理、欺诈检测、反洗钱和安全性领域。
- **社交网络分析。**探索社会实体之间的关系以及这些关系对个人行为的影响。这些功能可帮助将关系信息转化为表示个人和团体社交行为的关键绩效指标。

## 地理空间分析

借助 SPSS Modeler，用户能够探究可与位置绑定的数据元素之间的关系，并且对数据进行地理空间分析，以揭示无法从图表中直观看到的深刻洞察。借助空间数据挖掘功能，用户可以利用 ESRI 图形文件轻松挖掘地理空间数据。通过分析非空间与空间数据，可提高模型的整体精确度，用户能够对人员和事件有一个更深刻的认识。

揭示空间与非空间属性之间的关联规则，开辟一个全新的分析维度。通过空间时间预测功能，用户可以为在 2D 空间位置上获得的测量值拟合线性模型，从而轻松地预测“热点”区域以及这些区域随时间的变化情况。该技术可应用于诸如犯罪模式分析、疫情监测、房屋管理和分公司绩效分析等领域的地理空间数据挖掘工作。

## 文本分析

SPSS Modeler 的交互式可视化环境采用高级语言技术和自然语言处理 (NLP) 快速处理非结构化文本数据。SPSS Modeler 从该文本数据中提取关键概念并进行整理。借助可定制的行业专用文本分析程序包，除缩写、表情符号和俚语外，用户还可以在合适的上下文中分析相关术语和短语。交互式图表有助于探究与显示文本数据和模式，供即时分析使用。用户可创建层级化分类结构，并将其作为预测性模型的输入，以获得更合理、目标更突出的决策和成果。可导入诸如层级、注解和关键字描述符等预定义类别，对非结构化原始数据进行分类，这样用户可以用更符合逻辑的方式对概念进行更详细的整理。

## 灵活部署，满足个性化需求

在企业内部署分析方案取决于诸多环境因素，其中包括亟需处理的商业问题、操作系统与平台的选择以及企业基础架构所采用的其他技术和数据源。技术，尤其是软件，应当足够灵活，以适应各种排列并仍可提供预期的性能与效果。

SPSS Modeler 属于开放型架构，支持多种平台和语言。

用户可通过本地、云端甚至混合云部署 SPSS Modeler，然后放心地将其与原有系统配套使用，从而优化企业绩效并解决商业问题。这一灵活的部署方案通过按计划或需求为人员和流程提供结果，可填补分析与行动之间的空档。可出于模型刷新或作业调度自动化的目的部署 SPSS Modeler 串流。利用嵌入式预测性建模 (EPM) Java API，开发人员可将 SPSS Modeler 串流嵌入 Java 应用程序，而且可将这些串流与决策管理或其他预测性应用程序一起部署。

## 决策管理

决策管理将 SPSS Modeler 的预测功能延伸至日常业务流程中，为一线员工和系统提供有力支持。决策管理将预测性模型、简单规则与评分功能集成到企业系统中，方便管理和优化大量决策，并实现决策自动化。然后在必要时向人员提供行动建议，如对客户进行电话交叉销售、决定索赔的最佳途径、利用实用程序进行带宽分配或在自助服务亭进行产品推介时。大量的决策在操作层面即可作出，并且完全符合企业的目标和策略。

利用 SPSS Modeler 的决策管理功能：

- 预测性模型可预见最有可能出现的结果，识别促成结果的因素，如客户对特定产品的反应倾向，或特定索赔属于欺诈性索赔的风险。
- 商业规则使由诸如企业政策或法律法规合规性等元素确定的参数实现自动化。SPSS Modeler 直接提供基础规则支持。对于更多可扩展以满足企业级要求的稳健规则，也支持集成 IBM Operational Decision Management 方案。
- 综合评分功能可向合适的人员和系统即时提出合理建议，以便作出符合资源现状且遵循公司策略的决策，不论公司处于何种行业。

## 优化

即使是最热诚的企业也不可能给予每个客户最优待遇，保险公司也不可能调查每起索赔。商业运作必然受到实际条件的制约，包括可用人员、设备和投资方面的限制。通过识别最符合特定目标的解决方案，优化功能使企业可以对稀缺资源加以充分利用，获得最大的经济效益。示例包括：最大限度提高营销活动的收益，或最大限度降低诈骗或客户流失的风险。优化功能可按需提供，如针对个人，可确定最适合该个人的产品，或批量提供，可向所有合格客户分配产品。



### 执行与调度

SPSS Modeler 包括用于通过自动化提高经营成果一致性的功能。由于管理层能够高效管理分析流程开展的业务环境，经营成果的一致性可以增强人员对分析方案的信心。这种管理有助于确保企业内外的程序要求都得到遵守。

利用 SPSS Modeler，分析师可构建灵活并且可重复的分析流程，便于操作实施，也就是说，该流程可适时启动并且与企业其他流程整合。预定义模型管理流程有助于确保模型的相关性与准确性。

### 数据库内分析

SPSS Modeler 提供许多功能，可最大限度减少数据传送，将分析固定在数据库内，如：

- **SQL Pushback**。利用 SPSS Modeler Server，无需移动大型数据库，甚至是 IBM® System z® 和 IBM PureSystems® 中的数据，在数据库内即可进行数据分析与挖掘。SQL Pushback 功能支持数据库内数据转化与准备，无需编写任何 SQL 或进行任何编程。结果是分析性能显著提高。
- **数据库内评分**。数据库专用评分适配器适用于 IBM SPSS Modeler with Scoring Adapter for zEnterprise®、IBM DB2®、IBM PureData® System for Analytics（由 Netezza® 提供支持）和 Teradata 解决方案，可扩展可在数据库内进行评分的 SPSS Modeler 算法的数量，进一步减少评分前的数据提取工作。
- **数据库功能**。SPSS Modeler 可利用几乎所有用户定义函数 (UDF) 以及数据库提供的数据库聚合和窗口化聚合函数。这些功能显示在 SPSS Modeler 工作台上，用于扩展本地可用功能，并且确保 SQL Pushback 功能的正常运行。

- **数据库内挖掘**。SPSS Modeler Server 支持集成以下方案提供的数据挖掘功能、建模工具和数据库原生算法：IBM DB2 Analytics Accelerator (IDAA) on Z、“Hytap”（混合事务和分析处理）、PureDataSystem for Analytics、Oracle Data Miner 和 Microsoft Analysis Services 等。用户可利用 SPSS Modeler 工作台在数据库内建立模型、为模型评分并存储模型。

### 集成 IBM 技术

SPSS Modeler 支持将数据导出到 IBM Cognos® Business Intelligence 和 Cognos TM1® 软件中。分析结果可发送至仅需要结果的关键决策人，供汇报、监控和规划工作使用。需要进一步分析时，SPSS Modeler 也可作为数据源访问这些结果，这意味着可继续进行分析流程，从而将结果追溯至问题发生之处。

IBM SPSS Statistics 作为 SPSS Modeler 及其数据挖掘功能的补充功能，在各节点上具有专门划分，可进一步执行统计分析和数据管理。

SPSS Modeler 支持 PureData System for Analytics 通过 SPSS Modeler 界面访问特定模型并利用硬件的速度和性能。

对于大数据分析，在 SPSS Modeler 的基础上添加 SPSS Analytic Server，支持在分布式 Hadoop 中处理分析。

SPSS Modeler 通过 SPSS Analytic Server 与 Apache Spark 集成，可提供许多大数据算法和内存中处理功能，用于创建快速高效的机器学习模型。

SPSS Modeler 串流也可与 InfoSphere® Streams 一起部署，用于需要预测性评分的高速流数据应用场合。

## 结语

SPSS Modeler 是一款预测性分析平台，其范围从桌面部署到与操作系统集成不等，可提供智能预测，优化个人、团队、系统和企业决策。企业可利用 SPSS Modeler 进行数据分析，不论数据的位置、规模或类型（结构化或非结构化）如何。可扩展主从式架构支持用户访问从平面文件到大数据环境的一切内容。分析回归到数据源上进行，可最大限度减少数据传送量，同时提高分析性能。

各类用户均可利用 SPSS Modeler 解决许多商业问题。SPSS Modeler 提供从描述性分析到高级算法的多种分析技术，其中包括自动化建模、文本分析、实体分析、社交网络分析、决策管理与优化。直观的界面设计适合从非技术商业用户到专业分析人员的各类用户使用。SPSS Modeler 学习曲线短，对于新手和高级用户极具吸引力，他们很快就能发掘洞察，取得真正的商业成果。

## 了解更多

如需了解更多预测性分析软件和资源，请访问：  
[ibm.biz/predictive](http://ibm.biz/predictive)。

在本公司充满活力且不断发展的开源社区，用户可找到关于如何扩展 IBM 预测性分析软件用途的资源。这些资源包括博客、视频、教程，以及具有超过 6,000 项预测扩展程序的扩充程序库，可协助用户采用流行的编程语言，如 R、Python 和 Java。访问以下网址加入社区：  
[developer.ibm.com/predictiveanalytics](http://developer.ibm.com/predictiveanalytics)

## 关于 IBM Analytics

IBM Analytics 软件可提供数据驱动型洞察力，帮助企业以更加智慧的方式运营，并且超越竞争对手。这一全面的产品组合包括商业智能、预测分析和决策管理、绩效管理及风险管理解决方案。

借助 IBM Analytics 解决方案，如客户分析，企业可识别并直观地看到对企业绩效具有重大影响的行业趋势与模式。他们可比较场景，预测潜在的威胁与机遇，作出更佳规划，制定预算和预测资源，平衡风险与预期回报并努力满足监管要求。通过全面化分析，企业可以将策略和战略决策协调一致，以实现商业目标。如需更多信息，请访问 [ibm.com/analytics](http://ibm.com/analytics)

## 请求致电

要请求致电或提出问题，请访问 [ibm.com/analytics/contactus](http://ibm.com/analytics/contactus)。IBM 代表将在两个工作日内答复您的问题。



© IBM 公司版权所有 2016

IBM Analytics  
Route 100  
Somers, NY 10589

美国印制  
2016 年 2 月

IBM、IBM 徽标、ibm.com、Cognos、DB2、InfoSphere、PureData、PureSystems、SPSS、System z、TM1 和 zEnterprise 是国际商业机器公司在全球许多司法辖区的注册商标。其他产品和服务名称可能是 IBM 或其他公司的商标。当前的 IBM 商标列表请见网站的“版权和商标信息”版块：[www.ibm.com/legal/copytrade.shtml](http://www.ibm.com/legal/copytrade.shtml)。

Netezza 是 IBM 旗下公司 IBM International Group B.V. 的注册商标。

Microsoft 是 Microsoft Corporation 在美国和/或其他国家/地区的商标。

Java 和所有基于 Java 的商标和徽标为 Oracle 和/或其子公司的商标或注册商标。

本文档包含截至发布之日的最新信息，IBM 可能随时更改。本文档所述各种产品或服务不一定在 IBM 开展业务的所有国家/地区均有提供。

用户应自行负责使用 IBM 产品和程序评估和验证其他所有产品或程序的运行情况。本文所载信息按“原样”提供，不做任何明示或暗示的担保，包括对适销性、特定目的适用性的任何担保，以及针对非侵权的任何担保或条件。IBM 根据产品交付协议中规定的条款和条件为产品提供担保。



请回收再利用