

SPSS 数据挖掘系统

一个能帮您按计划节省时间和财力，实施数据挖掘项目的便捷指导。

目录表

简介	2
什么是数据挖掘？	2
数据挖掘与 OLAP、报告有什么不同？	2
数据挖掘和统计学有什么不同？	2
为什么采用数据挖掘？	3
数据挖掘用于解决哪些商务问题？	3
数据挖掘如何进行工作？	3
数据挖掘技术	4
如何成功的进行数据挖掘	4
跟随 CRI SP-DM 阶段	5
商业理解	5
数据理解	7
数据准备	8
建立模型	9
模型评估	11
成果发布	11
选择数据挖掘工具	12
有关 SPSS 公司	14
SPSS 产品	15
术语词汇	16

简介

您是否正在做数据挖掘项目？您是否正在考虑第一次承担一个数据挖掘项目？无论您是否有数据挖掘经验，SPSS 数据挖掘系统，都可以为您在计划和实施数据挖掘项目时，提供方便的指导。

在数据挖掘过程中应用 SPSS 数据挖掘软件，用它来节省金钱，及时实施您的项目并得到最好的结果。

如果您对开始或实施您的数据挖掘项目有疑问，请拨打 +86-01-62193608/62190468-8003。我们会提供不同的技术培训和咨询项目来帮助您。如果您对数据挖掘有任何建议或想法，可在线访问我们的网站：<http://www.spss.com.cn>

什么是数据挖掘？

数据挖掘是在大型数据中发现隐含模式和关系的过程。数据挖掘解决了普通的疑惑：你拥有越多的顾客信息，有效地分析和得出有意义的结论则越困难，越耗时间。数据有大量有价值的信息，却经常由于缺少才智，时间或技术，而未被开发。数据挖掘是用一个清晰的商务定向和强大的分析技术来快速、完全地挖掘山一样的数据，取出有价值的、有用的信息----你所需要的商务才智。

数据挖掘和 OLAP 和报告有什么不同？

报告和 OLAP 是用于理解过去所发生事务的重要工具。数据挖掘则是用于了解将来可能发生事务的方法。数据挖掘用预测性模型（包括统计和机器学习技术----例如神经网络）来预测将来。

例如，查询和报告告诉您上个月的总体销售情况。OLAP 则层层深入地告诉你上个月各项产品的销售情况。然而，数据挖掘会告诉你下个月谁可能买你的产品。而且，为了最好的商务效益，将开发和数据挖掘相结合，以发现如何使产品个性化以导致最大可能性的购买。

数据挖掘和统计学有什么不同？

数据挖掘并不是对统计学的代替。实际上，统计学是对数据挖掘的很好的补充。经典的统计学技术，如回归，与数据挖掘技术----如神经网络一起应用。统计学也可用于验证数据挖掘结论。

为什么要用数据挖掘？

当您有一个对您商务前景的可靠指导时，那您现在就有能力做出正确的决定。数据挖掘，通过了解过去和现在，得出准确的预测，使您有能力掌握和改变您公司的命运。例如，数据挖掘会告诉您哪些人可能会变成有利可图的消费者，哪些人是最可能带给您回报的。有了对未来的了解，您可以通过只为那些可能给您带来回报的，有价值的消费者提供您的服务来增加您的 ROI。您的决定是基于合理的商务才智，而不是基于本能的反应。这些决定会带来一致的结果，使您处于竞争的前列。

数据挖掘用于解决哪些商务问题？

您可以用数据挖掘来解决几乎所有与数据相关的商务问题，包括：

- 从消费者处增加收入
- 了解消费者的划分和优先顺序
- 识别有利可图的消费者，争取新的顾客
- 交叉销售和提升销售
- 保持消费者和提高忠实度
- 增加 ROI 和降低改进成本
- 检测欺骗、浪费和滥用
- 检测信贷风险
- 增加网络地址收益性
- 增加商店交易，最优化摆放产品以增加产品销售
- 追踪商业成绩

数据挖掘处理如何进行工作？

SPSS 数据挖掘产品和服务通过支持交叉行业数据挖掘标准流程（CRISP-DM）而保证及时、可靠的结果。由工业专家创造的 CRISP-DM 为数据挖掘过程的每一阶段的任务和目标提供一步步的指导。CRISP-DM 是工业标准化的数据挖掘过程。

CRISP-DM 阶段：

- 商业理解：明确了解您所面临的商务挑战
- 数据理解：决定什么数据可用于数据挖掘，以期得到答案
- 数据准备：以合适的格式来准备数据，回答商务问题
- 建立模型：设计数据模型来满足您的要求
- 模型评估：用您的结果逆向检测您的项目目标

■ 成果发布：使您的项目结果有用于决策者

CRISP-DM 是 CRISP-DM 社团合伙人的独享财产，NCR 系统工程，

数据挖掘技术情报处理系统

如何成功地进行数据挖掘

1. 跟随 CRISP-DM

在您的数据挖掘项目中应用 CRISP-DM 指导方针会起很大作用，它会引导一个成功的数据挖掘项目，而不是失望。跟随一个已证实的方法是很关键的----在商务问题中，复杂的数据挖掘技术和大量的可用数据会把没有稳固基础的项目压倒。

例如，没有专门的处理，你可能会掉进通常所说的陷阱：“让我们通过某些数据挖掘运算法则来处理数据，看看我们可以得到什么结果。” 结果是：数据模式可能并不适用于您的情况。

2. 以牢记结果而开始

为了能在项目结束时显示 ROI，在你开始前要了解你将要怎样评估结果（即，你应该用什么公司来测量，他们应该如何被计算？）。

例如，你是否想在 20% 的订户中找到 70% 的流失者？你是否知道，如何将信息译为，基于你的持续项目的成本和反应的假设下，收入提高期望值？或者，你是否知道，如果能辨认出 10% 的欺骗行为，你会节省多少钱？

3. 设置期望

确保项目股东知道数据挖掘并不是可以神奇地解决商务问题的智囊。数据挖掘只是一种在计算机支持帮助下的半自动的商务过程。有任何商务问题时，股东需要发现可解决的问题和想出解决办法。

如果你计划为你的市场部细分消费者，那么要让部内成员了解他们可能收到的信息----作为项目的结果----的种类。（即，我们正在应用产品信息和人口统计数据，所以我们期望得到基于年龄、收入等的细分，及显示这些消费者所喜欢的产品混合情况。）

4. 限制你初始计划的范围

以有限的目标和计划开始。当你达到成功时，再移向更复杂的计划。

例如，不要尝试立即提高每个地区的消费者获得量，总销售量，增长销售量和保持量，而是集中在更小、更实际的目标上。

5. 确认一个团队来完成数据挖掘

一个数据挖掘项目是需要集体努力的。数据挖掘要求商务用户了解问题和数据，以及懂得分析的人。也需要为数据拥有者提供入口。

例如，你需要一个数据挖掘分析者，一个数据库分析者，一个市场经

理。这些人可能会因项目目标不能站在一起，而掉进目标的不同功能领域。因此，寻找方法使他们的角色能合作良好是很重要的。

6. 避免数据垃圾

经常性的建立商务问题，定义项目目标和获得支持。如果你简单地提取没有项目结构的数据堆，你会在数据中迷失，而且浪费时间。不要让数据的量来牵动你的项目。要聚焦在商务目标上。你不可能利用你的所有数据，因为只有其中的一部分是与项目相关的。你甚至可能会发现你的数据对于解决你的商务问题来说是不充分的。大量的数据并不是你有正确数据的保证。

例如，最近的信息通常较大量的历史性数据能提供更准确的消费者行为的预测。

跟随 CRI SP-DM 阶段

这一章列出一些从数据挖掘指导----“CRI SP-DM. 1.0”----中摘录的提示。详细的工具在此处列出的信息处有扩展，并包含用户指导。可以在下面的网站中下载：

www.crisp-dm.org

商务理解

7. 从商务角度知道“谁，什么，什么时候，哪个地方，为什么和怎么样”

发展一个对项目参数的完全的理解：目前商务情况，此项目的基本商务目标，成功的标准和谁来决定项目的成功性。

8. 制定一个发展战略

考虑一下你想如何利用数据挖掘项目的结果。例如：

- 此结果是否将被不需要对结果进行翻译的专家所用？
- 此结果是否将被需要对结果进行不同程度翻译的广大雇员所用？
- 此结果是否将被需要某种格式的某一媒体（如在线，报纸等）所用？

9. 制定一个持续战略

一旦初始的结果完成后，你将怎样管理你的数据？如果项目在发展战略过程中，你将：

- 周期性的分析新的数据？
- 实时分析新数据？

10. 评估形势和详细目录资源

确保预先了解项目的每一方面以保证你拥有成功的必要条件：

- 个体（项目资助者，商务和技术专家）
- 数据来源（数据库或者操作数据）
- 计算机资源（硬件，平台）
- 软件（数据挖掘和其它相关软件）

11. 项目要求是什么？

下面列出了项目要求：

- 完成计划
- 结果的易理解和质量
- 安全性
- 数据应用的法律限制

12. 项目所做的假设是什么？

下面列出并澄清你已经有的所有的假设：

- 数据的质量（准确性和可利用性）
- 外部因素（经济论题，竞争性，技术领先性）
- 内部因素（商务问题）
- 模型（是否有必要向高级管理了解、描述或解释模型）

13. 项目操作的限制性因素是什么？

检查和发展下列问题的解决方案：

- 一般性限制（合法性问题，经费，时间，资源）
- 数据资源利用权（限制，必要密码）
- 数据的技术可利用性（操作系统，数据管理系统，文件或数据库格式）
- 相关知识的可获得性

14. 是否每个人讲相同的语言

确保每位参与者理解整个项目过程中所用的术语和想法。

通过制作与该项目相关或特殊的商务和技术术语词汇表，使部门间更容易地相互理解。

15. 将商务目标翻译为数据挖掘任务

决定你必须完成的数据挖掘任务以便达到你的商务目标。用技术术语对数据挖掘任务下定义。

例如，商务目标是：“增加现有消费者的目录销售量”可能被翻译为数据挖掘目标：“假定有前三年消费者的购买量，相关人口学信息和项目

价格，预测消费者会买多少小部件。”

16. 决定数据挖掘成功标准

运用技术术语，描述为了考虑项目成功而必须满足的标准。

例如，模型显示明确的预测准确性水平，或购买倾向性必须有一个明确的升高度。

17. 产生一个项目计划

制定一个计划，概述你将采取的每一步及要达到的数据挖掘目标和要满足的商务目标。评述要用什么工具和技术来使你完成计划。

数据理解

18. 确保数据是可利用的

收集本项目所需要的所有数据。如果你的数据有多个来源，确保你的数据挖掘工具能合并数据。

不同来源的数据可能产生质量问题（即，不同的属性或数值名）。关注潜在的问题。

约 80% 的数据在文本文件中可能被掩盖。用文本挖掘工具有效地搜索这些有价值的信息资源。

19. 对数据进行探索性分析

通过分析小量的多来源数据和交流发现，帮助数据库构造者设置优先权。这会有助于你确认重要的区域和潜在的自我实践区域。

20. 你的数据是否覆盖相关的特征？

通过选择最能代表欲分析的环境或行为的数据而确保成功。也决定你的发展选项是否能处理你所有的数据和限制你的数据的实际开发。

21. 描述现有的数据

通过产生一个描述数据格式、记录和区域数目、区域身份和其它相关特征的报告来得到你的数据的清晰图像。

22. 检查数据质量

为了防止出现问题，需要评估数据的质量和对被检测的任何问题作一计划。

- 属性的名字和他们所包含的值是否一致？
- 是否有属性被遗失？是否有任何空白区域？
- 检查值的多个拼写以避免重复。
- 寻找偏离和确定原因。

- ✓ 检查任何属性的应答是否与常理违背（例如，青少年有高的收入）
- ✓ 排除那些不相关的数据（即，如果检查家庭借贷行为，则要把没有家庭的人排除在外，等等）

23. 产生数据质量报告

检查数据副本，潜在的数据错误（如，在他们变成消费者前，显示他们已经流失）和可能包含错误信息的强制性的数据库区域。

数据准备

24. 选择您的数据

决定什么数据用于分析，并列出作此决定的原因。

这包括：

- 完成显著性和相关性检测来决定包括哪些字段
- 选择数据子集
- 用抽样技术检测小块数据的适当性。
 - ✓ 决定某些属性是否较其它的更为重要，并对他们进行加权处理。

25. 描述数据质量问题

为了确保可靠结果，现在花时间来调适任何数据问题。内容可能包括：

- 决定怎样处理噪声数据
- 指出特定值和它们的意义。有一个例子是，当一个问题没有被回答或者当数据由于空间的考虑而被截短时（例如，2002 可能会变为 02），可能会被默认为某一特定值。
- 某些字段可能与你的目标不相关，也不需要被清除。对那些字段采取跟踪行为或不跟踪，因为你可能会在后来的过程中决定用它们。

26. 选择一个灵活的数据重构工具

确保你选择的数据挖掘工具能够根据项目需要来构建数据库。你的工具应该允许你增加所需的新的字段。要记住数据挖掘是一项发现驱动型的过程----预先是不可能知道数据会带你走到哪里的。

27. 决定是否产生衍生属性？

由于下列原因，你可能希望产生衍生属性：

- 由于你对环境的便捷经验，你知道，尽管某一属性当前不存在，但它是非常重要的。
- 模型运算法则仅仅处理某种数据类型，因此，如果不重新生成变量的话，它不能被包括进去。

- 模型结果揭示：相关事实未被提出。
 - ✓ 在增加衍生变量前，决定它们是否或者如何对模型过程起到帮助作用。

28. 通过合并数据来巩固信息

当你加入新的表来巩固信息时，你可能也会想产生新的字段，汇总值。也可能要从非电子数据源（纸张报告，专家意见，等等）获得数据。

29. 你的数据挖掘工具是否要求特定的顺序？

在此阶段，如果你的数据挖掘工具要求你的数据集呈特定的顺序，你可能需要整理你的数据子集。

30. 数据是否应该被平衡？

检查你的模型技术是否需要平衡数据。

例如，直邮促销经常返回反应信息偏向“无反应”。某些技术作出无反应预测，因此有一个高的准确性。然而，为了准确地预测正反应，某些技术可能要求你使用大致相等的数字。

模型

31. 选择模型技术

为了将你的数据与正确技术相匹配，检查每种技术对数据格式和质量所作的假设。在某些情况下，只有一种技术适合于您的情况。确保要考虑：

- 什么技术适合于分析您的问题
- 是否有任何政治要求（管理期望，易懂性）？
- 是否有任何限制（特定的数据特征，职员技能，时间主题）？

32. 在构建前检测

在你产生你的模型前，检测你计划应用的技术的质量和正确性。产生一个包含训练测试，测试和确认的检测设计。然后构建训练的模型，用测试数据子集评估它的有效性。

33. 构建您的模型

为了产生一个模型，将您已经准备好的数据子集运行模型工具。描写结果，评估它预期的准确性，有效性和潜在的缺点。

- ✓ 产生一个详细的模型报告，报告中列出产生的规则，所用的参数设置，模型技术行为和解释，任何有关数据显示模式的结论。

在开发时，在正确场合下，用唯一的可用于模型的属性

例如，如果你想产生一个用于预测一年后消费者属性的模型，那么用从一年点到准确及时反映那一点的消费者行为的消费者数据。

用归纳法产生一个规则

规则是必要的参数，为了被考虑，数据必须落入参数中。

通常用“if/then”格式。归纳法使你能自动选择哪些规则对获得某一结果是最有效的。例如，用归纳法为贷款资格产生一套规则：

- 如果工作两年以上，则信贷风险是好的。
- 如果大于 30 岁，则信贷风险是好的。
- 如果在任何时候宣布破产，则信贷风险是差的。

34. 构建后检查

确保您的模型得出有助于你达到数据挖掘目标的结果。

- ✓ 用 lift and gain 表来检测模型的预测能力。

35. 试用几种模型以得到正确的拟合

为了改善模型性能，试着增加或删除项，或用可选项来实验。而且，虽然每项技术可能有轻微的差别，试着变化（如聚类和相关）来找到所有相关模式。

倾向模型对以下是适用的：

预测消费者行为----发现谁最可能购买，最可能？，最可能拖欠贷款和更多。用这些信息来决定哪些消费者和潜在的消费者提供最长期的利润。

聚类模型对以下是适用的：

发现有相似特征的案例的自然分组----运用聚类分析将异常的信用卡办理的相似案例分组而检测欺骗行为。

关联模型对以下是适用的：

市场篮分析----揭示哪些产品是最可能被一起购买的。利用这个信息可提高总的销售率通过分类和货架摆放，推荐火车，电话和直邮服务及其他更多。

Bonus tip from SPSS

统计模型对以下是适用的：

初始分析----统计分析在数据挖掘项目的早期是有用的，以获得数据结构的总体情况。一个简洁的数据特征的描述有助于项目成员得出假说，计划下一步的分析。

模型评估

36. 评估数据挖掘结果

检测通过给定模型得出的结论是否有助于您达到您的商务目标。是否因为商务原因而使模型有缺陷。

如果时间和资源是可利用的，试着检测模型或在实际应用中检测应用模型。

37. 为杰出完成任务回顾数据挖掘过程

当你已经证实了你的结果的质量和有效性后，回顾你的工作，以检查是否落掉了一些重要的步骤或信息。

- 是否数据挖掘过程的每一步在回顾中都是必要的？
- 是否每一步都是可能的？

38. 决定下一步

现在应该决定项目是否足够成功，可以向前开发。如果答案是否定的，做任何必要的步骤，以达到满意的结果。记住：

- 每一结果的潜在的开发
- 过程如何得到改善
- 是否存在一些资源，可以进行额外的步骤，或者重复先前的步骤

成果发布

39. 建立一个开发计划

取得项目结果，决定如何最好地利用他们来陈述你的商务论题：

- 总结可开发的模型或软件结果
- 开发和评估替代性开发计划
- 确保结果怎样被分发给接收者
- 决定如何监测结果的应用，测量效益
- 确定开发中可能存在的问题和缺陷

40. 监测和维持计划

确保你的数据挖掘结果得到最好的利用。可建立结果维持计划，描述：

- 未来中什么可能发生改变而影响结果的应用
- 如何监测结果的准确应用
- 如果必要，什么时候中断该结果的应用或开发

41. 建立最终报告

依赖于您的开发计划，报告可能是项目总结或是数据挖掘结果的最终陈述。为了建立最终报告，首先：

- 确定需要什么报告（幻灯片，管理总结等等）
- 分析数据挖掘目标如何被满足
- 确认报告接收人
- 列出报告的结构和内容
- 选择所要包括的发现

Bonus tip from SPSS

实施开发计划

根据开发计划分发您的数据挖掘结果，通过此而对结果选择性利用。如果不被用于改善商务，即使最聪明的发现也不会带来高的投资回报。

42. 回顾项目

这是一个机会，可评价什么正确，什么不正确，什么是主要成就，什么需要必要的改进。一个完全的回顾，试按下面进行：

- 回顾所有项目关键人员的经验
- 回顾任何数据挖掘结果利用者的经验
- 文件和分析您所做的个别的数据挖掘步
- 建立下一步计划的推荐

选择一个数据挖掘工具

本章中的提示是从 CRI SP-DM 文件：“实施数据挖掘工具评估”中摘录出来的。

43. 寻找一个已被证实的可用于解决您项目所陈述商务问题的数据挖掘工具

选择一种你所知道的，可以用于解决您公司问题，并在您计划应用方面有成功记录的工具。

44. 选择一种工具，用于在商务理解和数据挖掘技术方面起到沟通作用

确保工具所用的步骤与数据挖掘的商务需要相匹配：

- 工具是否可清晰地表达数据挖掘概念？
- 工具是否与项目管理软件或其它你可能用的工具相结合？如果不能，你是否不得不新建应用软件以弥补此不足。

45. 确保你的工具可对你现有的数据资源和格式进行操作

如果能选择一种能提取和合并多来源、多格式数据，你将会节省时间和金钱，并最大可能地得到可靠结果。这一点很重要，尤其是在数据挖掘过程的后期发现使你不得不从新的来源加入新的数据时。

46. 寻找交互式开发和可视化能力

选择一种可提供交互式可视技术的工具会使开发和理解数据变得容易。这些技术会使你通过在图内变化及根据不同的数据尺度产生新的图表更快地获得直觉。

47. 选择一种可高效，易解的进行数据准备的工具

选择一种可高效进行数据准备（从初始步骤到模型建立），且以易于理解的方式表达数据准备步骤的工具会节省时间和资源。这会使不同经验水平的项目成员获得有效的结果。

48. 确保你的工具可自动地提取数据

选择一种可为不同数据步骤自动提取数据的工具，可以避免耗时的人工书写查询。

49. 该工具是否可在合理的时间内建立有效的模型？

寻找一种工具，它可使分析家快速找到最有效的模型。这种工具应该支持有效的建立和检测多个模型。

50. 选择一种含宽范围技术的工具

为了确保最好的结果，确保你的工具能为可视化、分类、聚类、相关和回归提供一个宽范围的技术或运算法则。例如，你可能发现，对某一数据而言，一种技术比另一种更好。你需要能灵活地试用多种技术以获得准确，有效的结果。这种工具应该能联合应用在不同情况下可获最佳结果的多种技术。

51. 该工具是否可利用现有的数据和设备？

选择一种数据挖掘工具，它能利用现存的数据----或数据库中或文件中，也能与现有的分析和可视化工具相兼容。你不会愿意因为不能利用现有的数据库而浪费时间和资源再新建一个。

52. 选择一种可发送一致的、高质的结果的工具

要得到准确的结果，需利用在各种情形下都能很好工作的，适应性强的工具的数据挖掘环境和各种数据，而不是单一地为某一类型的数据或环境而设计的工具。您的工具应该能管理任何可能有助于阐释商务问题的数据。

53. 什么是工具的开发能力？

选择一种能将您的结果合并入现有的和将来的操作应用中的工具，是很重要的。因此也要考虑：

- 这种合并是否符合成本收益，或是否需要投入额外的时间和财力。
- 这种工具是否能容易地校正工具挖掘结果，如果可以的话，需要什么样的额外的投资？

54. 评估与工具相关的所有权潜在的成本

分析每项工具的潜在投资回报：

- 拥有产品寿命（包括工具所需的任何软件或服务）的成本是多少？你什么时候期望得到一个正的投资回报？
- 实施您的数据挖掘工具需要花多长时间？它是否为技术专家而设计？或者它是否适用于不同专业的用户？现在和将来的培训成本是什么？
- 这种工具是否为专为您的特殊用户和商务需要作了定制？您是否可节省普通的过程和使任务自动化？

关于 SPSS 公司

SPSS 公司，总部位于美国伊利诺斯州芝加哥市，是一个跨国计算机软件公司。提供运用预测分析和其它数据挖掘技术将数据转化为洞察力的技术。公司的解决方案和产品使组织者能通过对过去、现在的了解，及预测潜在的问题和机会而掌握未来。

要想了解更多的信息，请登录：www.spss.com.

SPSS 数据挖掘

SPSS 利用一个完全的数据挖掘方案----从理解您的商务问题到向决策者发放结果----可使数据挖掘过程呈流线和加速化。SPSS 利用已证实的 CRISP-DM 过程将数据挖掘能力带给各种水平的用户，所以你会得到更多的值和更好地利用结果。利用 SPSS 产品可达到广泛的商务目标：

- 增加商务单元和总利润率
- 理解顾客期望和需求
- 确认有利可图的消费者和获得新的消费者
- 保留消费者及增大忠实度
- 提高投资回报，降低提升成本
- 总销售和增长销售
- 测定信贷风险
- 提高网址收益性
- 提高商店交易量，为增加销售而使货架安排最优化
- 监测商务成绩

SPSS 产品

有了 SPSS 产品线，您可以建立一个可以满足您当今需求，并足够灵活可达到您未来目标的系统。

AnswerTree---- AnswerTree 可利用决策树来展示分割信息，预测群体如何反应。由于 AnswerTree 可显示可视化模型，你可以容易地看到有关群体。

Clementine — Clementine 交互式数据挖掘过程把您有价值的商务知识融入每一分析步骤，并建立强大的针对您的特定商务问题的预测模型。

Clementine Application Templates (CATs)---- 为顾客关系管理 (CRM)，网络挖掘和通讯交流的 CATs 可提供预建的普通数据挖掘应用软件数据流库，它能作为用户化数据流库的基础直接应用于您的数据或服务。

Clementine External Module Interface (CEMI)---- 通过增加你自己的程序，如数据处理程序或模型运算法则，使得 Clementine 用户化。

Clementine Solution Publisher---- Clementine 解决方案供应商可帮助你通过成本效益分析而做出更好的决策。

Lexi Quest™ Mine---- 利用 Lexi Quest™ Mine 来提示可能与您的数据挖掘项目有关的文档基础信息。Lexi Quest™ Mine 可快速地发现成千上万个资源，如文件、演讲、局域网和互联网的隐含信息。甚至能以每小时 25 万页的速度阅读 Lexi Quest Mine 可利用语言技术来辨认键概念和键间关系。

SmartViewer Web Server---- 有方便的，以网络为基础的 SPSS 报告发送，可得到为更好的决策而进行你所需要的分析。

SPSS Base---- 当你理解了数据时，你需要为分析而对它们进行准备。SPSS 是一个服务于分析过程----计划，数据收集，数据获取和管理，分析，报告和开发----的，有标准组件、紧密结合、全系列的产品线，也是数据挖掘程序的关键组件。

SPSS Advanced Models™---- 这一增强模板使你的分析更准确，提供更可靠的，含有适用于你的数据内部特征的程序设计的结论。

SPSS Business Intelligence Consulting---- 让我们的有经验的顾问帮你决定要解决的问题和如何解决。

SPSS Business Intelligence Training---- SPSS 商务智能分支可提供一个全套的数据挖掘课程和产品专业培训。大多数课程在 SPSS 简易版或你的公司位置即可使用。

SPSS 产品可用于 Windows™， Unix 和其它平台。

SPSS 有一个世界范围的发布网。有关 SPSS 产品的详细信息，请访问我们的网站：www.spss.com

专业词汇

Association 相关----发现同时发生的事件或有关事件的过程。例如，用相关技术来检测哪些产品经常被同时购买。与序列检测不同，序列检测用于检测产品被购买的顺序。

Balanced data 平衡数据----有时你会有两种或更多种数据要用来分析，每种数据应该有一个等值的数字使模型程序简化。

Churn 流失 -----流失描述顾客减少的过程，对许多公司，尤其是通讯业、财政服务业而言，这是一种主要的恶化来源。

Classification 分类----通过检测物体的特征而将其归类的过程。在分类时，群体被定义为一个外部标准（不同于聚类）。

Clustering 聚类----基于相似性而将记录分组的过程。例如，保险公司可能利用聚类分析，根据收入、年龄、购买的政策类型或先前的犯罪记录对顾客分组。聚类划分子集以使有相似内容的记录在同一组别内，而组与组之间尽可能不同（不同于分类）。

CRISP-DM ----CRISP-DM 提供了数据挖掘项目的结构，及潜在问题和解决方案的指导。CRISP-DM 分为六个阶段：商务理解，数据理解，数据准备，模型，评估和新产品发布。

Cross-selling——交叉销售向现有顾客提供或销售额外产品或服务的实践。

Data mining 数据挖掘----分析数据以发现隐藏模式和关系----有助于你管理和改进你的商业----的过程。

Decision trees 决策树----树状的图表显示清晰地显示数据分割、模式和等级。

Deployment 成果发布----数据挖掘结果的分发和应用。

Field 区域----一个单一数据或信息的区域，也称为变量和属性。例如，一个数据区域可能包含顾客的名。下一个数据区域可能包含顾客的姓。

Gains tables 利润表----利润表通过显示用模型所得结果和不用模型所得结果的差别，以显示模型效益。

Kohonen network Kohonen 网络----一种用于聚类分析的神经网络。

Lift charts 通过显示用模型所得结果和不用模型所得结果的比率使用户检测模型效益。Lift 线距离基线越远，模型越有效。

Machine-learning techniques 机器学习技术----一套可使计算机学习某种任务，如作决定，评估，分类或预测（无手工程程序）----的方法。

Model 模型----一套典型的法则、行为或特征，根据它们分析数据以找到相似性。描述性的模型被用于分析过去的事件。预测性模型用于提示未来

将要发生的事件。在预测性模型帮助下，数据挖掘者可发掘另外的想法以决定哪些行为会产生未来预想的结果。

Neural network 神经网络----一个利用复杂的、模仿脑细胞抽象功能的数学计划，用于预测或对个体进行分类的模型。通过用大量的观察个体来对神经网络进行训练，一次用一个，允许它不断地更新自己，直到它学会这项任务。

Noise 噪声----模型与它的预测之间的差别。有时，数据包含错误----如缺失或不正确数字或有一些外来列的混入----时，会被考虑为噪声。

OLAP 在线分析处理----在线分析处理使用户能分析当前的和历史的数据层。虽然在线分析处理可利用你的数据，告诉你正在发生的事件和已经发生的事件，但它不能告诉你未来将要发生的事件。

Pivot tables 枢轴表----交互表可使用户通过早期重新设置数据的行、列、层来得到信息的不同显示。

Predictive modeling 预测模型----建立模型以预测未来活动、行为或特征的过程。例如，一个预测模型可基于以前的流失者的特征和行为，来显示未来哪些顾客最可能流失。

Query 查询----基于某些特征或性质的信息的发向数据库的查询。

Record 记录----记录是指贮存在一起的一套相关数据。也称为行(电子表中)或个体(统计学中)。

Regression 回归----有许多类型的回归，包括 ordinal 回归，非线性逻辑思特回归，二项式逻辑思特回归，非线性回归(NLR)和非限制性非线性回归(CNLR)。

Reporting 报告----以一定为接收者易于理解的格式开发或发布数据分析结果的过程。

Return on investment (ROI) 投资回报率---- ROI 是被返回的或从各种在技术、下层构造等投资中获得的价值。

Rule induction technique 规则归纳技术----为预测或从样本个体中将未来个体分类的自动起源的决策决定规则的技术。