

现代预测性分析 业务指南



IBM

目录

为何使用本指南?	03
概览	04
为何预测性分析与 AI 非常重要	05
AI 采用率的转折点	06
AI 如何增强您的业务?	08
攀登 AI 阶梯	13
您有什么解决方案可选?	14
迈出下一步	16
要点总结	17
为什么要结合决策优化?	18
术语表	19



为何使用本指南？

在商业领域, 远见决定一切。如果能预测到接下来会发生什么, 就可以执行下列任务:

- 进行更明智的决策
- 加速进入市场
- 扰乱竞争对手

借助现代预测性分析的力量, 企业可以采用实时洞察来增强历史数据, 然后运用这些数据预测和塑造企业的未来。

预测性分析是分析发展历程上的一个重要里程碑。经典统计分析技术与人工智能 (AI) 新世界在这里交织汇聚。

[Forrester Research](#) 的研究结果显示, 广大企业已经到了开始将机器学习与知识工程相结合的时候。用人类智慧增强数据将显著加快 AI 应用的发展。

本指南将帮助企业实施以下行动:

- 纵览现代预测性分析技术的发展图景
- 发现促进和增强 AI 应用的机会
- 助力数据科学团队和业务利益相关者能够快速交付富有价值的成果

现代预测性分析旨在利用机器生成的预测结果, 结合人类本身具备的洞察力, 来积极推动业务发展。



概览

随着 AI 革命的兴起,企业越来越多地要求其数据科学团队解决一些重大问题。

因此,数据科学家们所要做的不仅仅限于一次性的研究项目。他们需要找到可重复、自动化的方法,为日常决策提供实时洞察。

为了满足这些需求,数据科学领导者不仅需要向业务利益相关者说明现代预测性分析技术的潜力,而且还需要确保成功交付结果。

定义并执行成功的数据科学战略的能力,将是未来几年里领导者和追随者之间的主要区别之一。

这项工作并不简单。培养数据科学能力将涉及以下活动:

- 吸引并留住一支高技能专家组成的多元团队
- 赋予他们无缝协作的能力
- 实施健全的治理架构,确保预测模型始终得到公司的信任

最重要的是,数据科学团队与业务团队需要找到全新方式进行有效合作。这些方法包括了解预测性分析的作用,以及确定 AI 在哪些方面有助于增益业务优势。

- 1 我们的客户表现如何?
- 2 市场为什么在波动?
- 3 是什么让我们的业务战略取得成功或遭遇失败?
- 4 接下来会发生什么?
- 5 项目资金从何而来?
- 6 购买中心在哪里?



为何预测性分析与 AI 非常重要

预测性分析并非新的概念。多年来, 统计学家一直在利用决策树、线性回归以及逻辑回归来帮助企业对数据进行关联和分类, 并据此做出预测。

不同之处在于, 预测性分析的范围已经拓宽。机器学习和深度学习领域的突破带来了许多机会, 使得企业能够在一些领域使用预测模型, 而在此之前, 这对大多数商业投资而言都是不切实际的。

许多企业看到, 直观易用的工具、新型预测技术以及混合云部署模式正在以前所未有的方式融合在一起, 这让预测性分析比以前更容易获取和使用。

因此, 这里就形成了一个转折点。有史以来第一次, 各种规模的企业或机构都可以进行以下活动:

- 将预测性分析内嵌至其业务流程中
- 大规模利用 AI
- 从之前未曾探索过的“黑暗数据”中提取价值, 包括原始文本和地理位置信息等一切内容

如果能从部门级小组 AI 项目演进到企业级数据科学平台, 您的企业或机构将获得显著的竞争优势。不把握该机会就可能落后。

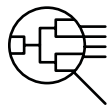
776 亿美元

截止到 2022 年, 花费在认知与 AI 系统上的金额

(数据来源: [IDC](#))



AI 采用率的转折点



哪些类型的数据可以分析？

之前：

主要是大规模的关系型数据；其他类型的数据需要特别的或专门的研究项目。

现在：

关系型数据、半结构化文档、文本、传感器数据等；可以大规模进行历史分析和实时分析。



有什么工具可用？

之前：

工具迥异，互不兼容，需要具备不同专业技能的多个团队之间多次交接。

现在：

拖放式界面和开源笔记本相结合，让团队之间的协作更加方便。



可以使用哪些分析技术？

之前：

基本统计技术，如逻辑回归与线性回归。

现在：

由先进机器学习和深度学习算法加持的统计技术。



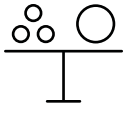
企业如何部署分析应用程序？

之前：

应用程序和分析技术绑定了本地数据服务器及数据仓库设备，减少了随时随地开展分析的机会。

现在：

混合多云部署有助于将分析技术应用于有数据驻留的任何地方，并在确保本地安全性的同时增加灵活性和可扩展性。



企业如何将分析集成到业务流程中？

之前：

生成静态报告供业务专家人工分析。

现在：

将预测模型无缝内嵌至新的应用程序和企业应用程序中。



企业如何实施治理规范？

之前：

仅在部门层面上临时性地遵循政策内容，可见性或可追溯水平极低。

现在：

连贯一致的治理规范和安全框架，能够在企业范围内大规模实施政策。



企业如何将人工智能注入到现代应用程序中？

之前：

应用程序开发团队和数据科学团队之间完全脱节，这意味着每一次部署都要规划定制内容。

现在：

数据科学生命周期旨在为 AI 集成创建一套标准化、可重复的流程。



企业如何在分析历程中取得进展？

之前：

从描述性分析到预测性分析，以及指导性分析，每一步都需要不同的工具、技能和投资。

现在：

由一个综合性平台来支持分析进度，简化引导流程，并随着需求的变化和技能的发展而不断演进。

AI 如何增强您的业务？

理论上讲，采用现代方法进行预测性分析就是一项简单直接的任务。技术不再是障碍，更好的工具大大降低了使用门槛。

但在实践中，交付富有价值的成果可能仍然是一大挑战。业务利益相关者尤其容易陷入有关 AI 的炒作中，对数据科学能够实现的目标抱有不切实际的期望。

定义用例

数据科学负责人和企业领导的首要任务是共同努力，确定现代预测性分析技术能够在哪些具体的实际用例中创造出富有价值的成果。

某些用例 (示例如下) 可能适用于大多数行业：

- 销售和营销团队的产品推荐以及“下一步最佳行动”模型
- 客户支持团队联络中心的自动化

其他用例可能特定适用于某个行业、部门乃至企业内的团队。这些用例往往更难执行，但它们更有潜力释放出独特的竞争优势。

哪些业务功能主导着企业在 AI 系统方面的投资？

46% 销售和营销

40% 客户支持

(数据来源: [Forrester Research](#))



一般用例

当一家企业开始投资一项新技术时,往往会着眼于最容易取得成果的地方,这很容易理解。

预测性分析也不例外。有些用例广泛适用于各个行业,并且供应商已经开发了通用的预打包模型和服务。

对于那些希望将数据科学从研究功能转变为日常运营的内嵌式组成部分的企业来说,这些服务应当是一个很好的起点。它们易于部署,只需极少量的定制开发,还能快速交付富有价值的成果。

联络中心优化

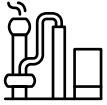
对许多客户服务团队而言,处理数量难以预测的客户电话、电子邮件、短信和聊天信息是一大挑战。

要减轻员工的压力,并减少客户等待时间,智能聊天机器人是一种强大且经济高效的方式。这些聊天机器人使用以下功能来了解客户咨询:

- AI 语音识别
- 自然语言处理
- 内容分析(以探索公司的知识库并找到有用的答案,无需人工干预)

现代预测性分析可以提供的一些最常见的跨行业用例,具体包括如下内容:

- 推动交叉销售和追加销售—通过个性化实时推荐和优惠活动
- 提高忠诚度—通过预测客户流失率,并进行预防性干预
- 优化产品和服务—通过倾听客户并预测未来需求
- 加强营销—通过开展有针对性的个性化推广活动
- 最大程度地降低库存成本并改善资源管理—通过准确预测
- 提高生产力—通过将正确的员工在正确的时间分配到正确的岗位上,并进行准确的劳动力预测
- 降低维护成本—通过预测故障,防患于未然
- 降低风险—通过实施准确的客户信用评级
- 检测欺诈—通过识别可疑的行为模式
- 开启新的商业模式—通过满足未开发的潜在需求,并将预测功能集成到现代应用程序中



适用特定行业的用例

在许多行业,为解决特定的业务问题,创新型企业已经在投资建立自己的预测模型。接下来的两页内容将重点介绍几大行业中 AI 和预测性分析的几种潜在应用。



保险

保险公司利用预测性分析执行以下工作:

- 检测欺诈性索赔
- 通过评估每个申请人的相关风险来优化报价和保费
- 预测危险天气事件以减少汽车保险索赔



商业银行业务

商业银行利用预测性分析来执行以下工作:

- 评估市场以及交易对手风险
- 评估贷款申请的信用风险
- 实时检测欺诈交易
- 利用预测建模来加快贷款审批流程



能源与公用事业

公用事业机构利用预测性分析执行以下工作:

- 管理庞大的实物资产网络
- 预测生产和需求模式
- 提前预测故障
- 供需规划



政府

政府依靠准确的统计数据,为许多领域的决策提供信息,包括以下预测性分析用例:

- 检测福利欺诈
- 预测公共服务的使用模式
- 优化废物管理和交通流量



制造

制造商利用预测性分析执行以下工作:

- 通过产品质量建模和缺陷检测,保持生产线的平稳运行
- 优化仓库管理和物流
- 利用机器学习模型开发自动驾驶汽车传感器



零售

零售商利用预测性分析执行以下工作:

- 管理客户忠诚度计划
- 根据客户资料和先进的倾向模型提出有针对性的建议,以此促进交叉销售和追加销售
- 实现准确的需求预测



食品

食品行业利用预测性分析执行以下工作:

- 自动收集和分析有关食品健康的数据
- 预测并警告潜在的卫生突发事件,以便快速干预
- 保护企业的敏感数据,确保竞争对手之间可以安全地开展合作



医疗保健

医疗保健机构可以利用统计建模技术执行以下工作：

- 监测来自心电图仪及其他医疗设备的数据流
- 预测患者的病情可能在何时发生变化
- 开展医学研究
- 实时分析患者数据流



零售银行业务

零售银行利用预测性分析执行以下工作：

- 通过加快信用评分提高客户满意度
- 通过混合云架构将灵活性与健全性及安全性结合起来
- 利用创新架构来降低成本并加快开发



运输

运输/物流公司利用预测性分析执行以下工作：

- 优化路线规划
- 实现车辆的预测性维护
- 优化供应链运作

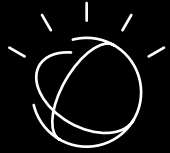


教育

教育机构利用预测性分析执行以下工作：

- 预测学生的成绩和续读率
- 确定需要额外支持才能实现目标的学生
- 加强捐助者关系
- 跟踪学生动向, 帮助减少缺勤现象

攀登 AI 阶梯



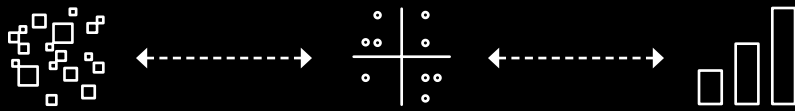
注入 - 利用信任和透明度进行 AI 运维活动

分析 - 利用 AI 在任何地方扩展洞察能力

组织 - 创建可信的分析基础

收集 - 使数据简单易用

各种类型的数据, 无论位于何处



通过现代预测性分析取得成功都是一个长期的历程。重要的一点是将 AI 战略定位于企业的正确层级, 并将技术和组织成熟度纳入考量。数据科学负责人和企业领导需要共同努力, 才能确定创造商业价值的最佳、最快的方式。

从技术角度看, 可以将 AI 成熟度想象成一个阶梯。阶梯的第一层是数据收集, 因为, 如果没有数据, 就没有任何东西用以分析或建模。第二层是数据整理。添加用于治理规范及可发现性的元数据, 确保所需数据始终可供数据科学家使用。

虽然数据的收集和整理是重要主题, 但不在本指南的范畴内。我们是要重点探讨如何帮助组织登上阶梯的以下两个最高层级:

- 通过建立、训练和测试预测模型来分析数据
- 通过将模型作为应用的一部分部署至生产中, 从而将 AI 注入运营活动

您有什么解决方案可选？



IBM 的 AI 产品组合可为企业提供登上 AI 阶梯顶端所需的一切。

[Watson Assistant](#) 和 [Watson Visual Recognition](#) 等预置 AI 服务可帮助企业快速高效地完成常见用例，从而快速创造富有价值的成果。

当企业准备开发自己的 AI 解决方案时，[Watson Studio](#) 和 [Watson Machine Learning](#) 可以为预测模型的构建、训练和部署提供无缝工作流程。这些解决方案通过利用最先进的 IBM 工具和最好的开源 AI 框架来增强企业的实力。

[Watson Knowledge Catalog](#) 针对模型和数据提供强大的数据治理和可发现性功能，而 [Watson OpenScale](#) 则帮助企业实时监测和管理模型，从而提高准确性，增强可解释性，并减少偏见的发生。

[IBM Cloud™ Pak for Data](#) 统一了所有这些功能的访问渠道，并提供了一个强大的多云数据平台。

IBM Cloud Pak for Data 专用的 IBM Data Science Premium 附加组件可提供额外的数据科学生产力，如 SPSS Modeler 和 Decision Optimization，以加速实现价值成果，并增加 AI/ML 项目取得成功的几率。



AI 阶梯的最高两个层级是“分析”和“注入”。为了达到这两个层级，企业或机构必须帮助数据科学家和业务利益相关者在数据科学生命周期的各个阶段有效开展合作。

整个生命周期由以下三个相互作用的子周期组成：

构建

数据科学家通过探索业务数据来确定需要关注的特征，然后着手准备那些用于设计预测模型的精良数据集。

运行

运维团队对模型进行训练、测试、部署和管理，并在必要时进行重新训练。

管理

业务专家监测模型的运行时性能；寻找任何偏见迹象，或确定是否需要说明；提供反馈；并在需要重新训练时通知数据科学团队。

迈出下一步

根据自身在 AI 阶梯上的进展水平,企业可能会根据自己组织内预测性分析采纳的水平而有不同的要求。

启动

当企业开始培养其数据科学能力时,他们通常从特别的项目着手,开发模型来回答特定的问题或支持一些研究项目。通过 Watson Studio Desktop 等解决方案,数据科学家可以在自己的计算机或笔记本电脑上 7x24 小时地工作,并在需要时与更广泛的团队同步动态。

成长

当数据科学被广泛采用时,不同的部门都需要部署各自的模型,与数据源建立连接,并将其注入到生产类应用程序中。通过 Watson Studio 和 Watson Machine Learning,部门级的数据科学和 IT 团队可以在这整个生命周期中更轻松地开展协作。

应用到整个企业

一旦 AI 被内嵌至关键业务流程中,构建一个中央平台对于模型和数据的管理和治理将至关重要。IBM Cloud Pak for Data 可以为发挥单个控制点作用的综合性多云平台提供所需的架构和工具。

求真务实

无论您是一位数据科学家还是企业领导,如需了解 IBM 的现代预测性分析产品组合在业务转型中发挥的具体作用,最好的方式就是亲身体验。请试用以下某个教程来开始体验:

进行机器学习练习

在 IBM Watson Studio 中使用 Apache SystemML 进行练习,以此深入了解机器学习。[了解更多信息](#)

创建一个评分模型来预测心率衰竭

使用 IBM Watson Studio,构建用 IBM Watson Machine Learning 的预测模型。[了解更多信息](#)

使用物联网 (IoT) 传感器数据预测设备故障

了解 IBM Watson Studio 如何分析多变量物联网传感器数据并预测设备故障。[了解更多信息](#)

分析开放的医学数据集以获得洞察力

利用 IBM Watson Studio 运行机器学习分类器,并将输出结果与评估指标进行比较。[了解更多信息](#)

塑造并细化原始数据

使用 IBM Data Refinery,为预测性分析准备大型数据集。[了解更多信息](#)

要点总结

IBM 的现代预测性分析产品组合具有以下优点, 数据科学负责人和企业领导可以利用这些优点在 AI 时代牢牢把握竞争优势:

规模

- 通过自动化数据科学和数据工程任务来减少运维工作量和成本
- 在多种企业应用程序间无缝训练、测试和部署模型
- 将普通的数据科学功能扩展到不同的混合多云环境中

速度

- 利用预置的应用程序和预先训练好的模型来加速开发
- 帮助数据科学和业务团队开展协作, 从而更快地交付富有价值的成果
- 将 IBM 最先进的软件与开源软件相结合, 简化模型构建

简约

- 利用中央平台来管理整个数据科学生命周期
- 标准化开发和部署流程
- 在整个企业内建立单一的数据治理与安全框架

Watson Studio 帮助企业专心解决问题并发现机会。

[了解更多信息](#)

在 Watson Machine Learning 的支持下, 企业可以部署和管理模型, 以快速提供所需的结果。

[了解更多信息](#)

为什么要将决策优化与预测性分析结合起来？

IBM Decision Optimization 是一种指导性的分析解决方案，数据高度密集型行业可以通过它解决复杂的优化问题，从而做出更好的决策并实现业务目标。企业领导使用此工具来改善他们对资源的有效利用，包括但不限于以下活动：

- 供应链库存流
- 劳动力调度
- 运输路线规划

该解决方案利用机器学习类应用程序的预测结果来提供经优化的结果，从而与预测性分析实现良好的配合。机器学习基于用户的观察数据提供对未来的洞察。利用机器学习得到答案后，就可以训练机器如何找到该答案。

决策优化让您可以采取下一步行动，并根据这些信息行动。对于决策优化，虽然您不知道答案，但您非常清楚什么是好答案，什么是坏答案。拿到机器学习所输出的内容后，就可以指定要进行决策优化的操作，其中可能包括实现业务目标所需要的优化规则和限制。

完成该操作后，决策优化将返回一些能为企业交付价值的答案，例如可操作的项目和调整建议。通过该活动，决策优化将增强预测性分析所能发挥的作用。

通过该解决方案，团队可以将优化技术、机器学习技术与模型管理、部署及其他数据科学能力结合起来，开发出能够提高运维效率的最佳解决方案。

[了解更多信息](#)



术语表

算法是定义一系列操作的规则集, 这些操作可应用于数据, 以解决特定的问题。在数据科学背景下, 该术语囊括了大量的技术, 包括:

- 决策树与回归模型
- 自回归移动平均 (ARMA)、自回归综合移动平均 (ARIMA) 以及指数平滑法
- 带预测和异常值检测功能的传递函数
- 集成模型与层级模型
- 向量机与时间因果建模
- 用于时空预测的时间序列和空间 AR
- 生成式对抗网络 (GAN) 与强化

您的数据科学平台应当做到能让您轻松访问所有这些强大的算法。

人工智能 (AI) 是指计算机系统解释数据并从数据学习的能力。该术语最常用于描述那些利用机器学习或深度学习模型构建的系统。利用 AI 技术, 计算机可以解决之前被认为难以解决的各种问题。

偏见是在设计、训练和测试那些可能导致不准确预测的模型时比较常见的一类问题。随着企业越来越广泛地寻求采用 AI, 在运行时通过监测和审计模型来减少偏见正成为日益重要的主题。

分类模型旨在将数据点与已经分类的一组数据点进行比较, 从而实现了对数据点的分类。结果会得到一个离散值, 即有限选项列表中的一项, 而不是一个得分。例如, 对于客户是否有可能购买, 或者他们是否存在不良信用风险, 分类模型可以给出一个肯定或否定的答案。分类模型可以使用各种技术来构建, 包括决策树和逻辑回归。

内容分析是对各种格式文档中的非结构化数据进行分析, 包括文本、图像、音频和视频文件。机器学习技术可以大幅提升对大型内容存储库的分析速度, 而在以前, 这些内容需要花费工作人员数百或数千小时来进行审查和分类。

数据科学是一门所涉范围极广的学科, 它将统计学、数据分析和机器学习的各个方面结合起来, 利用数据来解决业务问题。

深度学习是机器学习的一个分支, 会用到具有大量隐藏层的神经网络。这些高度复杂的网络被用于深度学习的各种前沿领域, 如计算机视觉、机器翻译和语音识别。

训练深度神经网络所涉及的计算量非常大, 通常需要具有高性能处理器的设备集群。IBM Watson Studio 或 IBM Cloud Pak for Data 等混合云平台可以让各种规模的企业更容易负担和使用这种架构。

部署是将模型集成到业务应用程序中,并对照真实数据运行该模型的过程。要在测试、预发布和生产环境中制作和移动模型,需要数据科学团队、应用程序开发人员和 IT 运营团队相互协作。

将开源数据科学工具与企业现有的持续集成和部署管道整合起来会颇具挑战性。为了避免因人工部署而造成团队之间多次交接的情况,具备自动化部署功能、连贯一致的数据科学平台可能成为一项重要优势。

开发预测模型需要用到传统的统计技术或机器学习算法,创建好模型后,对照企业的数据集对模型进行训练和测试,以此来完善模型。

开发过程具有高度迭代性;可能需要训练几十甚至几百个模型才能达到要求的精度。这就是为什么让围绕模型开发和训练的工作流程实现自动化,就可以创造巨大的价值成果。

可解释性是任何利用预测模型来提出建议并协助业务决策之系统的一个重要属性。看起来复杂而神秘的预测模型可能很难令业务利益相关者、监管机构和客户相信其输出成果。Watson OpenScale 先进的运行时监测和日志记录功能可针对每个决策提供背景信息,使 AI 模型具有透明性和可审计性。

数据探索是模型构建过程的一个重要组成部分。该活动旨在揭示既定数据集中值得关注的特征,发现隐藏的关系,并重点关注那些通过预测建模可交付业务价值的用例。

在探索阶段,要运用数据科学技能和业务知识来定义您希望回答的问题以及希望预测的结果,这一点至关重要。这可能会导致准备和探索环节的迭代循环,直到您完成了对该领域的探索,并拥有了继续下一步所需的合理数据。

地理空间分析是对地理数据的分析,如纬度和经度、邮政编码和地址。这种分析对于解决许多实际的数据科学问题非常有用。现代数据科学平台应使地理空间信息的检测、解析和计算变得容易,并实现与绘图工具的轻松集成,以可视化方式呈现结果。

人工智能中的推理将逻辑规则应用于知识库,在存在不确定性的情况下得出结论。通过推理,用户可以得到经过简化、压缩和优化的运行时性能预测。

线性回归是利用一个自变量来解释或预测某个值或得分的统计过程。例如,在某一周内所售产品的 SKU 数量或客户关闭其帐户的风险百分比。

逻辑回归是一种用于预测结果的统计过程。该过程与线性回归的不同之处在于,一个自变量的可能值在数量上是有限的,并非具备无限的可能。当响应分为第一、第二、第三等以数字顺序整理的类别时,用户所使用的就是逻辑回归。

机器学习使用统计技术从大型数据集中导出复杂的预测模型和算法,而无需显式编程。

通常,在这个迭代过程刚开始的时候,您需要将数据集划分为两个子集,以便进行训练和测试。您需要对照训练集来训练模型,对照测试集来测试它们的性能,然后用几十甚至几百个变量来评估其预测准确性。通过执行这个过程,并根据每次迭代中性能最好的模型来生成下一代变量,如此,模型便会逐渐学习并提高性能。

机器学习的主要方法可分为两类:监督式学习和无监督学习。

模型管理对于确保模型长期保持准确性而言至关重要。定期对模型进行重新训练以纳入新数据的做法至关重要,因此,模型的开发、实施、部署和管理应当形成连续的循环。

使用不同的开源工具很难实现这种管理。使用端到端数据科学平台可以避免在这个过程中出现漏洞。该平台还可以确保相关团队及时得到通知,并在模型性能开始下降时迅速采取行动。

自然语言处理 (NLP) 是 AI 的一个领域,其主要目的是使计算机具备分析非结构化文本数据的能力。常见的用例包括语音识别、自然语言理解和情感分析。

神经网络为训练模型提供了一个框架,使许多机器学习算法之间能够进行复杂的交互,以帮助识别最优模型。

人类及其他动物大脑中有相互连接的神经元,这些神经元结构为人工神经网络结构的构建带来了启发。人工神经元由不同的层级连接在一起。在这个结构中,数据从输入层穿过一个或多个隐藏层到达输出层。在这个穿行过程中,数学函数将数据转换为预测,您可以对预测的准确性进行评估。

开源软件已在统计建模和机器学习的许多领域里成为了一种日益占据主导地位的范围。R、Python 和 Scala 等语言,Apache Hadoop 和 Spark 等大数据架构,以及 TensorFlow 和 Spark MLlib 等机器学习框架,都是预测性分析和数据科学领域的主要参与者。

开源框架通常侧重于开发高质量的工具,此类工具以数据科学流程的特定部分为目标,如模型开发或训练。因此,这通常需要让最终用户负责将所有工具集成到一套连贯一致的工作流程中。当您试图将预测性分析扩展到整个企业内,并将 AI 内嵌至业务流程时,这个任务就可能面临困难。

预测性分析使用历史数据对特定的领域或问题进行建模,并将过去促成了特定结果的关键因素分离出来。使用该流程建立的模型可以从新数据中预测未来可能的结果。

从经典的统计建模到机器学习算法,预测性分析所涵盖的技术极为广泛。

预测模型是将输入项(即一段数据,如数据库记录、文本样本或图像)映射到输出项或预测的算法。输出项通常是连续变量,如数字或百分比,或离散类别,如“是”或“否”。预测模型有两大类型:回归模型和分类模型。

数据准备是数据科学流程的第一步。大多数项目都从细化数据集开始,以确保数据质量足够好,能够承受详细分析。

在许多情况下,可能需要清理源数据,并将其转换为更易于建模和分析的格式。如果您正在构建一个机器学习模型,您可能还需要花精力以人工方式标记数据,以便在监督式学习中使用。

回归模型当您有一个包含多个变量的数据集并且希望分析它们之间的关系时会很有用。具体而言,当其他变量发生变化时,回归模型可以揭示某个特定变量可能会如何发生变化。

线性回归可用来预测一个值或得分。例如,在某一周内所售产品的SKU数量或客户关闭其帐户的风险百分比。

统计建模是一个数学领域,涉及到根据一组数据的概率性假设来创建模型。广大企业已经在使用统计模型来分析其数据集的重要特征,并确定可用来分类数据或生成预测的关联性。

监督式学习是一种使用数据集训练机器学习模型的方法,所用数据已经被正确标记。该模型会生成一个输出变量(通常是一个类别或一个值),因此,通过比较输出和标记的输入,可以轻松评估其准确性。线性回归、随机森林和支持矢量机都是监督式学习算法的常见例子,大多数预测模型都是使用此类技术建立。

测试预测模型对于确定 AI 流程中数据的准确性以及训练而言至关重要。预测模型需要不断测试来提高准确性。如果模型失败,分析人员必须找出根本原因,并进行重新训练和测试来改善模型。

文本分析利用语言规则、自然语言处理和机器学习来考量非结构化内容。该过程审查数据的方式与人脑审查数据的方式相同,但速度更快。通过文本分析,您可以从非结构化内容中获得更多洞察见解和发现,非结构化内容约占所有数据的 90%。

训练预测模型是机器学习、深度学习和其他 AI 流程的一个关键要素,用于确定哪些数据有用。在提供准确预测方面,经过训练的模型可用于对实时数据进行评分。为适应不断变化的行为模式,必须定期对模型进行重新训练。

无监督学习是一种使用未标记数据训练机器学习模型的方法。其目的通常是对数据中值得关注的模式或结构进行建模和强调。聚类和关联问题是非结构化学习的常见领域,例如,找到新颖的方法来细分客户或识别客户之间的相似度。

可视化是以图形方式表示数据的过程,通常会用到图表。为了理解数据,人们需要能以可视化方式呈现数据。在向业务利益相关者展示结果时,以及在项目的早期阶段探索新数据集时,这个流程都十分重要。

企业的预测性分析平台应当具备直观的图形界面和可视化工具。这些功能可以帮助用户在几分钟内开始理解那些哪怕是最大的数据集。



© Copyright IBM Corporation 2019

IBM Corporation
New Orchard Road
Armonk, NY 10504

美国出品
2019 年 3 月

IBM、IBM 徽标、ibm.com、IBM Cloud、IBM SPSS Modeler 以及 IBM Watson 是 International Business Machines Corp. 在世界各地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。其他产品和服务名称可能是 IBM 或其他公司的商标。以下 Web 站点上的“Copyright and trademark information”部分中包含了 IBM 商标的最新列表：www.ibm.com/legal/copytrade.shtml。

本文档为自最初公布日期起的最新版本，IBM 可能随时对其进行更改。IBM 并不一定在开展业务的所有国家或地区提供所有产品或服务。

此处讨论的性能数据系在特定操作环境下得出。实际结果可能有所差异。本文档内的信息“按现状”提供，不附有任何种类的（无论是明示的还是默示的）保证，包括不附有关于适销性、适用于某种特定用途的任何保证以及非侵权的任何保证或条件。IBM 产品根据其提供时所依据的协议条款和条件获得保证。

客户有责任确保其遵守适用的法律法规。IBM 不提供法律建议或表述或保证其服务或产品会确保客户符合法律法规的规定。

良好安全实践声明：IT 系统安全涉及通过预防和检测来自企业内部和外部的不正当访问并做出相应反应来保护系统和信息。不正当访问可导致信息被更改、破坏、盗用或滥用，也可能导致系统被损坏或滥用（包括用于攻击他人）。任何 IT 系统或产品都不应被视为完全安全，任何一个产品、服务或安全措施都不能完全有效防止不正当使用或访问。IBM 系统、产品和服务旨在成为合法、全面的安全措施的一部分，这必然涉及其他操作程序，可能需要借助其他系统、产品或服务才能发挥最大作用。IBM 不保证任何系统、产品或服务可免于或使您的企业免于受到任何一方恶意或非法行为的影响。