

数据科学和机器学习平台的魔力象限

2020 年 2 月 11 日发布-ID G00 385005- 阅读 84 分钟

从事数据科学工作的专家数据科学家和其他专业人员需要具备以下功能：数据源，构建模型和操作机器学习见解。供应商的显着增长，产品开发和无数竞争愿景反映了一个健康且迅速成熟的市场。



市场定义/说明

该魔力象限评估了数据科学和机器学习（DSML）平台的供应商。Gartner 将 DSML 平台定义为核心产品和相干集成产品，组件，库和框架（包括专有，合作伙伴和开源）的支持产品组合。它的主要用户是数据科学专业人员。这些人员包括专家数据科学家，公民数据科学家，数据工程师和机器学习（ML）工程师/专家。

一致的集成意味着核心产品和支持产品组合提供一致的“外观”，并创建用户体验，其中所有组件都可以合理地互操作以支持分析渠道。

DSML 平台提供了基本功能和高级功能的混合，这些功能对于构建 DSML 解决方案至关重要（主要是预测模型和规范模型）。该平台还支持将这些解决方案合

并到业务流程，周围的基础架构，产品和应用程序中。它在数据和分析管道中的多个任务中支持各种技能的数据科学家，包括以下所有领域：

- 数据提取
- 资料准备
- 数据探索
- 特征工程
- 模型创建和培训
- 模型测试
- 部署方式
- 监控方式
- 保养
- 合作

并非所有组织都从头开始或完全自己构建 DSML 模型。有些人需要帮助来开始或扩展他们的 DSML 计划。尽管此“魔力象限”确实评估了某些预打包内容（例如模板和示例）的可用性，但并未评估可帮助在整个组织中启动或扩展 DSML 项目的服务提供商。此“魔力象限”也未评估针对行业，领域或功能的解决方案的专业供应商。

魔力象限的读者应了解以下内容：

- 这个市场具有众多供应商：Gartner 邀请了许多 DSML 平台供应商参与评估过程，以将其潜在地包含在魔力象限中。这些平台的用户对用户界面（UI）和工具有不同的要求和偏好（请参阅“如何选择正确的数据科学和机器学习平台”）。专家数据科学家更喜欢对数据科学模型进行编码（通常使用 Python 或 R），并且经常在笔记本中构建和运行数据模型。其他用户最喜欢通过使用拖放 UI 来创建可视管道来构建模型。新兴的公民数据科学社区的许多成员都喜欢一种更加增强的方法（通常称为 AutoML）。它使用人工智能/机器学习技术来指导和加速从业人员通过模型构建和运营过程（请参阅“增强型机器学习正在使数据科学民主化”）。一些专家数据科学家更喜欢使用增强方法，这种方法使他们能够更有效，更高效地导航模型构建和操作过程。
- 明确排除了主要针对应用程序开发人员和业务分析师的平台：数年来，数据科学和传统分析空间一直在多个关键点发生冲突。此外，许多开发人员正在接受有关机器学习的教育，并将其纳入他们的应用程序和软件。角色和工具包之间的界线继续模糊。但是，在基于预期用户类型的市场空间之间仍然存在明显的区别。有

关为这些支持者设计的平台和服务，请参阅“面向云 AI 开发人员服务的魔力象限”（即将发布）和“面向分析和 BI 商业智能平台的魔力象限”。

- 领导者可能不是最佳选择：有许多可用的 DSML 产品，所有这些产品都具有广度和深度的功能，并提供了开发，操作和管理模型的多种方法。因此，在评估供应商时评估您的特定需求很重要。例如，“领导者”象限中的供应商可能不是您的最佳选择。同样，利基播放器可能是完美的选择。有关每个平台的功能的广泛回顾，请参阅（即将发布）随附的“数据科学和机器学习平台的关键功能”。请记住，此“魔力象限”仅包括提供各种 DSML 解决方案的众多供应商中的一小部分。
- 仅包括具有商业许可产品的供应商：纯开放源代码平台不包含在此“魔力象限”中。仅包含商业许可的开源平台。但是，我们确实认识到利用开源组件和库的商业平台的良好趋势。供应商采用不同的方法来包含和支持开源。开源解决方案为用户和供应商提供了一个机会，使他们只需很少的前期投资即可开始使用 DSML（请参阅注 1）。在开源社区中，创新是快节奏的，而新技术的成功是高度民主的。此外，许多 DSML 平台用户已经精通或可以轻松学习和应用开源技术。开源技术在大学数据科学课程中也变得无处不在。通过与商业产品的协作或精心集成来利用开放源代码，也减少了供应商重新创建特定功能的需求。供应商可以融合快速变化的算法和技术领域中的最佳元素，从而留出更多资源来专注于其平台的其他差异化领域。但是，如果平台的供应商未考虑所有类型用户的需求，则其易用性和一致性可能会受到影响。
- 平台不仅必须支持模型构建，而且还必须支持模型可操作性：除非同时具有以下两种模型，否则将无法实现 DSML 的全部收益（包括业务价值）：
 - 嵌入业务流程和决策环境中
 - 随着时间的推移进行维护，监控和管理
- 最近在技术，流程和人才开发方面取得了许多进步。但是，以全面部署为目的开发的模型的惊人数量从未真正投入运营。造成这种情况的原因很多，但一个关键的原因是缺乏实现和促进运营的工具。运营（通常称为 MLOps）扩展到对模型的持续审查和调整，以确保随着时间的推移，随着业务及其目标的变化，它们的相关性。MLOps 还包括关键功能，例如漂移检测，目录，治理，可解释性和业务影响分析。
- AI 被夸大且不可避免：所有 DSML 都可以归类为人工智能（AI），但并非所有 AI 概念都应称为 DSML。尽管如此，DSML 平台仍不可避免地被围绕 AI 的持续

炒作所淹没。这些语义不太可能达成共识，因此不值得争夺。AI 的炒作无疑给数据科学领域带来了宝贵的关注和热情。但是，如果没有教育，纪律和合理的期望，炒作的弊大于利。魔力象限中的供应商在宣传方面做得很好，同时清楚地交流和传递了价值，并使他们的平台与其他 AI 解决方案区分开来。

DSML 平台的多样性很大程度上反映了使用它们的人群的广泛性。因此，此“魔力象限”针对的受众是：

- **专家数据科学家：**这些人是备受追捧的人，他们具备理解和参与数据科学生命周期各个阶段的技能和知识。大多数专家数据科学家将大部分时间和精力花费在模型创建上，并由诸如数据工程师和 ML 工程师等支持角色来承担数据管道和 MLOps 职责。终身专家可以担任数据科学经理的角色，可以使用平台来了解团队的全部项目组合，并促进协作和及时交付价值。一些专家数据科学家大多独立于“点”解决方案上工作，很少与组织内的其他数据科学家或部门进行过多合作。
- **公民数据科学家：**越来越多的公民数据科学家正在构建 DSML 模型。这些人需要访问 DSML 功能，但不具备专家数据科学家的高级技能。公民数据科学家可以来自业务分析师，业务线 (LOB) 分析师，数据工程师和应用程序开发人员等职位。他们需要了解 DSML 市场的性质以及它与分析和商业智能 (BI) 市场有何不同，但又相互补充 (请参阅“分析和商业智能平台的魔力象限”)。公民数据科学家不会代替专家数据科学家，而是与他们合作。
- **支持角色：**这些角色包括数据工程师，开发人员，机器学习工程师和其他角色。尽管不负责模型的建立，培训和测试，但数据科学团队中的支持对于扩展操作，确保数据质量和一致的模型准确性至关重要。
- **业务线 (LOB) 数据科学团队：**通常，它们由其 LOB 执行官赞助，负责解决 LOB 主导的营销，销售，财务和 R&D 等方面的计划。这些团队专注于自己和部门的优先事项。与其他 LOB 数据科学团队的协作级别有所不同。LOB 数据科学团队可以包括专家和公民数据科学家。支持角色可以驻留在 LOB 中，也可以从 IT 或其他领域分配。
- **企业数据科学团队：**这些团队拥有强大而广泛的执行官赞助，并且可以从企业范围内的可见性角度采取跨职能的观点。除了支持模型构建外，他们还经常负责定义和支持用于构建和部署 DSML 模型的端到端流程。他们通常与多层组织中的 LOB 数据科学团队合作。此外，他们可能会为没有自己的数据科学家的 LOB 团队提供帮助。公司数据科学团队通常包括专家数据科学家。支持角色可以驻留在公司数据科学团队中，也可以从 IT 或其他领域分配。

魔力象限

图 1. 数据科学和机器学习平台的魔力象限



资料来源：Gartner (2020 年 2 月)

供应商优势和注意事项

牵牛星

[Altair](#) 的总部位于美国密歇根州的特洛伊.2018 年 12 月, Altair 完成了对 Datawatch 的收购, Datawatch 于去年 1 月收购了 Angoss 及其主要数据科学产品组件。Datawatch 产品组合现在称为 Altair Knowledge Works, 而本“魔

力象限”评估考虑的核心产品是 Knowledge Studio。Knowledge Works 套件还包括用于 Apache Spark 的 Knowledge Studio，知识搜索者，知识中心，知识管理器，Panopticon 和 Monarch。

Knowledge Works 的在线用户社区使客户能够使用这套产品与最终用户和数据科学专业人员进行协作和学习。为此，Altair 在多个城市提供了一系列的客户培训日。

除了正常的客户支持外，还可以协商价格提供现场和高级支持。Altair 还提供各种咨询和托管服务产品，以支持构建和部署模型。

Altair 作为利基市场参与者进入了魔力象限（Datawatch 和 Angoss 都是前两年的利基市场参与者）。其产品远景完整性和执行能力的得分因该产品套件多年来进行的第二次收购相关的风险和不确定性而受到影响。Altair 可能是不断发展的 Datawatch 和 Angoss 组合投资组合的理想之家，但过渡期仍在进行中。Knowledge Studio 在供应商活跃度高的几个领域也落后，包括增强的 DSML，MLOps 和尖端用例。

长处

- 编码器和非编码器的易用性和吸引力之间的平衡：Knowledge Studio 高度评价的 UI 将吸引面向视觉的数据科学专业人员。编码人员还将找到一个灵活的环境，在其中可以充分利用 Python 和 R 的优势。大多数现代数据科学团队都需要这种双重吸引力。该界面有机会通过添加增强的 DSML 功能来扩大和增强其吸引力。
- 客户满意度：Altair 凭借 Knowledge Works 继承了客户成功的悠久传统。借助 Altair，客户可以跨多个接触点感到满意。这包括帐户管理，集成和部署以及常规服务和支持。
- 新产业的发展潜力和其他 Altair 产品的发展潜力：知识工程将继续以其传统优势（决策树，金融服务用例）而闻名。与 Altair 产品的集成和整体策略将打开更多的机会。Altair 能够将 Knowledge Works 引入新的市场和销售渠道。

注意事项

- 两年内第二次收购产品组合：无论是战略上还是燕尾形的所有收购都带有不确定性和风险。考虑到过渡期，Altair 不能延迟客户和潜在客户要求的重要更新和路线图交付。Altair 面临的挑战是如何继续整合和优化对 Datawatch 的收购，同时在一个竞争激烈的，往往是残酷的市场中定位和推进 Knowledge Works。
- 愿景和当前产品之间的差距：Altair 需要在关键领域(例如增强型 DSML, MLOps 和可解释的 AI) 赶上思想领袖。需要对容器化和 Kubernetes 的更强大支持。Altair 将需要在这些方面进行改进，以保持专家的吸引力和公民数据科学家的思想份额。
- 与竞争对手相比，采用速度较慢：与竞争对手的产品相比，核心知识产品的市场采用速度很慢。特别是在过去的两年中，尽管 Knowledge Works 易手，但一些竞争对手却保持了高速增长，并开发了市场领先的产品。

Alteryx

[Alteryx](#) 的总部位于美国加利福尼亚的尔湾。它提供了四个构成其 DSML 平台的软件产品：Alteryx Connect, Alteryx Designer, Alteryx Server 和 Alteryx Promote。Alteryx Designer 是核心产品。

Alteryx 托管一个 Alteryx 社区，该社区为客户，合作伙伴和 Alteryx 员工提供了一个可以相互交互的平台。Alteryx 每年在北美，欧洲和澳大利亚组织 Alteryx 启发会议。该供应商为美国，英国，澳大利亚和新加坡办事处以外的客户提供多语言的全球支持。

Alteryx 通过展示扎实的公司和产品愿景，特别是在增强 DSML 和流程自动化方面的表现，已从 2018 年成为挑战者，重新回到 2018 年的 Leader 职位。从收入增长的角度来看，该供应商在该魔力象限中的表现将继续超过几乎所有其他供应商，并且已在国际上显着扩展了业务。

Alteryx 在改变其仅是数据准备提供者的观念方面取得了重大进展。它在 2019 年进行了两次战略收购 (ClearStory Data 和 Feature Labs)，以扩展其平台功能。ClearStory Data 提供了一种解决方案，可在诸如 Apache Spark 之类的大规模数据处理平台上自动化对复杂数据和非结构化数据的分析。Feature Labs 使功能工程，AI 应用程序的创建和数据准备过程自动化，以帮助提高模型准确

性和总体过程效率。Feature Labs 的收购发生在本研究报告的截止日期之后，因此并未计入评分。

长处

- 市场感知和执行：Alteryx 专注于解决端到端分析过程，这显然将其定位为具有全面平台的供应商。市场不再将其简单地视为数据准备供应商。Alteryx 的价值主张在市场上清晰明确并引起共鸣，其强劲的收入和客户增长证明了这一点。
- 无代码方法：Alteryx 的无代码方法对于希望使用 ML 但需要为业务分析师和公民数据科学家构建的易于使用的平台的组织很有吸引力。该供应商为用户提供了数百个自动化构建块，以快速构建分析，数据科学和流程自动化 workflow。该平台可以直接与机器人流程自动化 (RPA) 解决方案和企业应用程序集成，以自动化 ML 管道和业务流程的运行。
- 一个强大的，不断发展的社区：Alteryx 建立了一个参与的在线用户社区，以分享知识和回答问题。它的在线论坛提供挑战练习，以帮助用户提高其产品知识。此外，它还提供了三个快速发展的年度用户大会（美国，欧洲和亚太地区）。

注意事项

- 定价：由于市场上存在开源和其他低成本选项，Alteryx 的高价继续引起一些客户的关注，这可能会限制整个组织的广泛使用。一些客户报告了许可的复杂性，并由指定用户和 CPU 内核收取费用。批量折扣的缺乏还限制了组织内部平台的广泛使用。
- Linux 支持：在四种 Alteryx 产品中，只有 Promote 可以在 Linux OS 上运行；而在 Linux OS 上只能运行 Promote。CentOS 是唯一受支持的发行版。Alteryx Connect 和 Server 仅在 Microsoft Windows Server OS 上运行。对于不支持 Windows Server 的组织，这可能是一个问题。
- 流/物联网 (IoT) 功能：Alteryx 支持流/物联网的功能有限，这对实时决策至关重要。示例用例包括与物联网和物流相关的用例，例如故障预测，资产管理和供应链优化。

水蟒

[Anaconda](#) 位于美国德克萨斯州奥斯汀市。Anaconda 提供 Anaconda Enterprise (此分析不包括 Anaconda 发行版)，这是一种基于交互式笔记本概念的数据科学开发环境，支持用户使用基于 Python 和 R 的开源软件包。Anaconda 继续提供一个松散耦合的分发环境，该环境可以访问主要基于 Python 的各种开源软件包和开源库。

该公司致力于建立一个庞大而活跃的开源和企业用户社区，其需求是 Anaconda 创新的重要推动力。AnacondaCON 是供应商的年度会议。活跃的社区是用户支持的重要来源，尽管企业客户也可以依靠支持台和专业服务。

Anaconda 在今年的“魔力象限”中仍是小众玩家。它非常适合使用 Python 或 R 并且渴望探索连续不断的新开源功能，同时仍比纯笔记本环境更具结构性和集成性的环境的数据科学家。但是，这种灵活性是以缺乏自动化为代价的，并且要在频繁变化的开放源代码功能范围中寻找方法的复杂性。

最近，Anaconda 进一步改善了其可伸缩性，以服务于更多的并发用户。其他值得注意的改进是对安全性和图形处理单元（GPU）部署的更好支持。

长处

- Python 和开源支持：Python 在数据科学家中的流行使 Anaconda 对开发人员具有极大的知名度。Anaconda 为用户提供了对 Python，R 和其他开源工具，库以及快速集成创新的统一访问。
- 活跃的社区：Anaconda 是唯一的数据科学供应商，不仅支持而且还保护 Python 开源社区。该社区培育了先进的 Python 代码库，并与其他开源数据科学项目集成。Anaconda Cloud 为数据科学家和开发人员提供了协作和代码库交换的手段，以探索和加快模型开发的生产。
- 开源库的可伸缩性开发：Anaconda 的可伸缩性有几种形式。用户可以使用 GPU 扩展工作负载，并且可以在本地或云中利用 Apache Hadoop, Apache Hadoop YARN 或 Kubernetes 集群。

注意事项

- 专为专家设计: Anaconda 的目标读者是经验丰富的数据科学家, 他们通常熟悉 Python, R 和 Jupyter 等笔记本电脑。尽管这些环境具有灵活性, 但它们不利于与业务利益相关者进行富有成效的讨论, 而这种能力越来越受到希望改善业务与数据科学之间一致性的组织的重视。
- 开放源代码的缺点: 与许多开放源代码启动器一样, Anaconda 也遭受着与大型, 灵活的开发者社区相关的常见缺陷: 软件包和版本之间的兼容性问题; 新的, 有时是重叠的项目的不确定结果; 尽管在工作台同质性方面不断取得进步, 但总体连贯性仍面临挑战, 特别是对于非专业用户而言。
- 自动化和可操作性: Anaconda 环境在诸如 autoML, MLOps 和可解释性等功能方面落后。新手 Anaconda 用户将很难通过 Python “丛林” 找到自己的道路。公民数据科学家将发现自己处于 Anaconda 环境中的未知领域。

数据块

[Databricks](#) 的总部位于美国加利福尼亚州的旧金山。供应商的统一数据分析平台涵盖了数据科学, 机器学习和数据工程, 强调了基础架构自动化和可扩展性的简便性。Databricks 植根于 Apache Spark, 通过领导 MLflow, Delta Lake 和 Koalas 不断为开源社区做出贡献。其商业平台进一步使客户能够跟上流行的开源框架和库的各种发行版和快节奏的更新, Databricks 在 ML Runtime 下对其进行打包和优化。

Databricks 在美国和欧洲举行了两次 Spark + AI 峰会, 并在全球范围内举行了不计其数的聚会, 一直在领导 Apache Spark 社区。该公司不断发展其培训和认证计划。它拥有一个全球支持团队, 并根据部署规模和所需的响应时间提供几层支持。

在过去的一年中, Databricks 在产品开发, 收入增长和客户满意度方面表现出色, 尤其是在企业领域, 以及在欧洲和亚太地区的国际扩张。参考客户表示, 他们选择 Databricks 平台时的关键细节是供应商的产品路线图和未来愿景。这两个因素巩固了 Databricks 在此魔力象限中的领导者地位。

长处

- 强大的执行力，扩展能力和合作伙伴关系：Databricks 在多个行业和地区展现出强大的执行力并因此而增长。其垂直销售策略对其客户有效。Databricks 的增长还得益于一项完善的战略，该战略旨在开发一个由全球 500 多家公司组成的合作伙伴生态系统。
- 对客户成功的承诺：Databricks 旨在通过咨询服务，专业服务和培训迅速从平台交付价值。参考客户特别赞赏与客户并肩工作的客户成功工程师。
- Databricks 不断扩展可扩展性：Databricks 的平台可实现 ML 模型部署，基于笔记本的 ML 工作流，作业自动化以及数据基础结构和管道管理的可扩展性。参考客户特别赞扬其轻松且可扩展的集群管理。

注意事项

- 对技术受众的强烈偏爱：Databricks 向其核心受众传递的信息是技术性的。供应商依赖开发人员关系和拥护者。专有的数据科学笔记本解决方案 Databricks Notebook 是数据科学家在平台上开发解决方案的主要方法。对于具有编码背景的数据科学家和数据工程师，Databricks 统一数据分析平台的进入门槛相对较低。但是，对于业务分析师和其他新兴公民数据科学家而言，采用难度更大。
- 没有通常可用的本地产品：通常仅在选定的公共云中提供 Databricks 产品。可以从供应商处以一次性部署的方式采购本地 Databricks 平台。作为开放数据平台产品的一部分，目前也可以从第三方（该公司的 OEM 合作伙伴 Booz Allen Hamilton）处获得。
- 与 Spark 的紧密联系使 Databricks 不再是 DSML 平台：许多客户选择 Databricks 并不是因为他们正在寻找 DSML 平台，而是因为他们对执行 Apache Spark 的选项感兴趣。这些客户在 Amazon Web Services 或 Microsoft Azure 上拥有 Hadoop 或数据湖，他们将 Databricks 视为数据处理平台，而 ML 在其中。对 Spark 不感兴趣的公司通常不会将 Databricks 视为 DSML 平台选项。

达蒂库

[Dataiku](#) 的总部位于美国纽约，并在法国巴黎设有总部。Dataiku 的核心产品是 Data Science Studio (DSS)，它专注于跨学科的协作和易用性。

Dataiku 拥有一个在线用户社区，该社区在美国和欧洲的一些年度会议上开会。与合作伙伴一起，该社区可以通过市场提供和共享插件，以扩展 Dataiku 的标准产品。Dataiku 已将其联盟，合作伙伴和转售商的生态系统发展为全球 80 多个组织。它根据特定于许可证的 SLA 提供支持，以及针对策略，开发，实施和培训的专业服务。

Dataiku 今年被定位为领导者。这主要是由于它的易用性，远见和支持多种协作类型（从数据工程师和数据科学家到公民数据科学家）的能力。在 2019 年，Dataiku 继续增强了与可伸缩性，数据治理和增强型 ML 相关的功能。它已成为一个成熟的平台，支持各种用例。Dataiku 可以通过扩展其先进的实时分析功能来支持 IoT 和决策自动化用例，从而进一步提高其地位。

长处

- 跨数据科学角色的协作：从成立之初，团队合作一直是 Dataiku DSS 平台的核心。开发 DSML 模型涉及的所有主要角色在平台上都有其位置。DSS 提供了各种 UI 端点，所有端点均指向同一执行核心，这使其成为本“魔力象限”中评估的最一致的产品之一。
- 易用性：客户最常强调的 Dataiku 的品质是其平台相对易于学习，并提供了快速的生产力途径。增强的 ML 功能以及提供可解释或“白盒”模型的明确可能性，也有助于平台的直观感觉和短暂的学习曲线。然而，专家数据科学家也可以笔记本方式进行开发。
- 数据治理：在过去的一年中，DSS 扩展了其能力，以支持治理以及对 GDPR 和其他法律法规的遵守。这包括数据假名化，突出显示敏感数据以提醒用户数据保护规则和过程以及对敏感项目和数据源的限制性访问。框架方法允许纳入多个内部和外部政策和法规。

注意事项

- 先进的分析：Dataiku 改进了对图形和时间序列分析的支持，但似乎没有优先考虑仿真和优化功能，这在决策自动化的环境中变得越来越重要。宣布支持规则处理，但尚不可用。
- 预设的解决方案：Dataiku 不提供横向或纵向解决方案，而是依靠合作伙伴或其他第三方提供的插件。尽管供应商希望专注于开发和提供通用平台以进一步使数据科学民主化是有道理的，但 Dataiku 应该继续进一步加强其合作伙伴生态系统，以提供其解决方案来丰富其产品，以促进数据科学的采用。
- 流媒体和物联网：Dataiku 的 DSS 似乎受到以服务为中心的组织的青睐，并且在较小程度上受到资产密集型行业的青睐。Dataiku 不提供强大的 IoT 功能和实时分析功能（通常与以资产为中心的用例相关联）的事实可以解释这一点。

数据机器人

[DataRobot](#) 总部位于美国马萨诸塞州波士顿，提供增强的 DSML 平台，在端到端分析过程中整合了自动化功能。该平台使公民数据科学家可以有效地访问和使用数据科学技术，而专家数据科学家则可以提高效率，使用新算法以及测试偏差。

DataRobot 提供的 AI 成功计划通过现场工程师，AI 成功经理，客户主管和面向客户的数据科学家（CFDS）团队实现售前和售后支持。该团队与客户一起确定项目并确定优先级，并提供指导，支持和培训，以使模型投入生产并随时间进行维护。

DataRobot 广泛的合作伙伴关系和集成网络不断壮大，2019 年 10 月启动了一个新的用户社区站点，以促进用户与 DataRobot 社区之间的通信（在将其纳入魔力象限评估之后）。供应商在全球各地举行年度区域用户会议。

DataRobot 保持了自己的远见卓识。它于 2019 年 2 月收购了数据协作平台 Cursor; ParallelM, MLOps 平台，2019 年 6 月; 以及数据准备提供商 Paxata 于 2019 年 12 月（在纳入魔力象限评估的截止日期之后）。结合 2017 年收购 Nutonian，以在 2018 年增强时间序列建模和 Nexosis，DataRobot 继续为各种用户建立其端到端功能。

长处

- 增强分析功能可领导和吸引广泛的用户：DataRobot 负责将增强分析功能纳入 DSML。它继续定义和演示增强分析的使用，以与传统角色协作来吸引新型用户。DataRobot 提供了有益于多种角色的功能，包括开发人员，数据科学家，统计学家和业务分析师。

- 不断增加的市场吸引力：DataRobot 凭借收入，用户数量和全球知名度的显著增长，在市场上保持了稳固的吸引力。这归因于有效的市场营销和销售，以及通过 CFDS 和产品内协助提供的出色持续支持。
- 动手入门和支持：AI 成功计划及其相关团队可帮助用户发展内部数据科学能力，识别并确定项目优先级，快速启动计划并快速获得洞察力。DataRobot 还可以快速添加产品增强功能并响应客户要求。

注意事项

- 被视为非专业人士的入门基础工具：一些客户表示担心 DataRobot 解决复杂模型的能力。随着需求的成熟和增长，它们已经考虑到其他平台，并且在企业范围内也越来越多。DataRobot 需要改变这种市场观念，因为 DSML 和分析以及 BI 市场中的增强模型构建正在迅速变得商品化。
- AI 成功计划方法占用大量资源：尽管 CFDS 的使用效果很好，但是如果增长保持在当前的高水平并且客户不能从对 CFDS 的依赖中迅速过渡过来，该方法将难以维持。DataRobot 正在扩展其 AI 成功计划，但它还必须更加关注其使客户自给自足的意图，尤其是考虑到该平台易于使用的增强方法。
- 成本问题：虽然大多数客户参考都表明 DataRobot 具有良好的性价比，但人们认为成本仍然很高。许可成本使 DataRobot 不再需要某些组织考虑。成本问题也限制了某些客户更广泛地部署平台的能力。

骨牌

[Domino](#) 位于美国加利福尼亚州旧金山，Domino 是工业级，功能丰富且与工具无关的平台，用于云或本地端到端 DSML。在具有大型，松散耦合的数据科学社区的企业中，Domino 可以用作（并且已经用作）中央 DSML 平台。它可以容纳各种工具，并提供透明，可扩展的协作环境。

由 Domino 领导的年度会议 (Rev) 面向数据科学领导者和团队，在 2019 年的出席人数增加了一倍。该供应商在支持和服务方面的参考客户评分是所有评估的最高供应商之一。Domino 旨在帮助数据科学团队充分利用其软件，并使客户在 Domino 平台的部署，监视和使用方面变得自给自足。

Domino 在此魔力象限中被定位为有远见的专家，这主要是由于其产品的进步和发展路线图，这两者都表明了供应商对 DSML 市场的深刻理解。Domino 成功地使企业实现了其行业实力的部署。展望未来，供应商需要继续增长，并改善其营销，市场知名度和国际影响力。

长处

- 企业 ML 生态系统的中心：Domino 的开放式体系结构有助于在单个平台上整合所有数据科学资产和工作负载。该平台真正帮助数据科学团队协调和简化了 ML 工作流程，从环境管理到可重复性和最终部署，包括进一步的监控。
- 业务支持和协作：参考客户选择 Domino 来改善业务流程结果并支持业务用户与数据科学团队之间的协作。高度协作的环境使数据科学团队之间可以进行有价值的交流，并允许管理人员和业务利益相关者跟踪结果和进度。商业用户许可证是免费的。Domino 提供业务绩效跟踪，以显示 API，应用程序和其他已发布资产的端点参与情况。
- 治理和审计：治理和审计功能增强了 Domino 作为企业级平台的能力。其新的实验跟踪功能全面，对受监管的行业尤其有吸引力。Domino 可以通过一个受管框架在内部和外部各方之间协调 ML 和数据处理管道。

注意事项

- 伙伴关系策略：Domino 与一些云提供商和系统集成商建立了良好的伙伴关系。尽管如此，更大，更广泛的合作伙伴网络仍可以将产品交付给需要它的公司，从而提高供应商在 DSML 市场中的知名度，并有助于推动其财务成功。
- Domino 适用于大型团队：Domino 的协作，实验管理和其他企业功能可以很好地服务于大型团队，但是对于只有几个数据科学家的团队而言，这是多余的。数据科学计划不够成熟的公司可能不会立即受益于面向数据科学社区的平台，而不是单个数据科学家。
- 不灵活的定价：Domino 的定价和合同谈判抑制了产品的使用。供应商的许可模式需要更大的灵活性，以适应企业内各种实施和扩展方案。这对于 Domino 平台最具吸引力的大型分布式企业尤其重要。

谷歌

[Google](#) 总部位于美国加利福尼亚州山景城，提供 Google Cloud AI 平台作为其核心 DSML 平台。该平台具有扩展的产品套件，其中包括 Cloud AutoML，BigQuery ML 和 TensorFlow。

Google 拥有业内最大的 ML 堆栈之一。其中包括硬件，一流的深度学习框架（TensorFlow），强大的 AI API 和庞大的云执行基础架构。凭借对 Kubeflow 的承诺，它越来越多地在“随处运行”范例中执行。

但是，Google Cloud AI 平台当前不是独立产品。对于完整的端到端 ML 管道，客户需要使用多种其他 Google Cloud 产品（例如 Cloud Data Fusion，Data

Studio, Datalab, Deep Learning Containers, BigQuery, Anthos) 和非 Google 产品 (例如, Kubernetes 和 Trifacta 的 Cloud Dataprep) 。

Google Cloud 十分重视为开发人员, 数据科学家和 ML 专家提供支持, 尤其是在最前沿。它提供了大量的视频教程, 前沿的聚会和会议 (例如 Google Cloud Next) 。AI 枢纽和 Kaggle 社区进一步促进了协作, 该社区为 350 万用户提供服务。Google Cloud 与包括数百个系统集成商在内的众多合作伙伴建立了伙伴关系, 并且正在扩展其 ML 服务集。

Google 在这个魔力象限中被定位为有远见的人。尽管对于顶尖的数据科学人才来说, 这是一个绝佳的选择, 但 Google Cloud 仍在提高公民数据科学家的能力。展望未来, 使平台变得更加协调一致, 并为不那么先进的数据科学团队提供更易于访问的平台, 应该是当务之急。

长处

- 强调最前沿的数据科学问题: 对于开发人员, Google Cloud AI 平台提供了一个庞大的执行框架 Google Cloud Platform (GCP) 以及许多不同的路径来开发 ML 模型 (Cloud AutoML, TensorFlow, BigQuery ML, Kubeflow, 笔记本以及各种开源选项) 。Google 拥有丰富的 AI API 产品组合 (从文本, 音频和图像到视频处理) 以及众多用于大规模数据处理和集成的一流软件包 (例如 BigQuery 和 Data Fusion) 。
- 创新: Google Cloud 的 AI 平台提供高端可扩展性和内置的数据标记功能。它的 AI Hub (仍处于公开测试版) 是全球算法市场最有前途的实现之一, 可促进协作和重用能力。AI Hub 可以在内部充当模型存储库以及功能和管道存储。Google Cloud 还为 Cloud Dataproc 和 Cloud Dataflow 的流数据提供了强大的支持。
- 可扩展性: Google Cloud 的 AI 平台提供了可管理的云基础架构, 用于自动扩展从中型到超大型的不同硬件配置。它包括硬件和云加速器, 例如 Google Kubernetes Engine (GKE) 集群自动缩放和 Cloud Tensor Processing Unit (TPU) Pod。

注意事项

- 庞大的足迹和学习曲线: 使用 Google Cloud 的 AI 平台创建完整的 ML 开发管道可能具有挑战性, 对于缺少 GCP 基础架构知识的中小型数据科学团队而言尤其如此。但是, 即使对于有经验的 GCP 开发人员, AI 平台工具链也包含多个相互连接的组件, 并且学习曲线很大。

- 机器学习流水线仍然不连贯，缺乏项目管理支持：工具链尚未完全适应数据科学项目的需求。取而代之的是，Google Cloud 工具用于许多其他行业目的。重用部分实验管道很困难。Google 的路线图建议在 AI Hub 的未来版本中解决这些问题。
- 永久性 Beta 和本地功能有限：像去年的评估一样，许多 Google Cloud 最具差异性的功能仍处于公开 Beta (例如 AI Hub, AI Platform Data Labeling Service, Cloud AutoML Tables) 。而且，某些创新组件 (如 AutoML 和所有 AI API) 仍无法在本地使用。

过氧化氢

H2O.ai 位于美国加利福尼亚的芒廷维尤 (Mountain View) ，它提供名为 H2O 无人驾驶 AI 的商业产品，并支持开源产品 H2O-3 (在与 Apache Spark 集成的版本中称为 Sparkling Water) 。此外，还有一个称为 H2O4GPU 的开源 GPU 加速 ML 包。供应商的优势包括其高性能 ML 和对重要趋势 (如增强数据科学和可解释性) 的愿景。

H2O.ai 提供全球客户支持，并辅以社区闲散渠道。为了提供建议和咨询，H2O.ai 提供了访问其专家数据科学家和 Kaggle Grandmasters 的权限，这是所有商业活动的一部分。H2O.ai 有广泛的合作伙伴，并已开始与其中一些合作伙伴一起开发针对特定业务领域或行业的首个解决方案。

H2O.ai 保持了去年魔力象限中有远见者的地位。在所有评估的供应商中，它的“愿景完整性”是最强的，而由于在欧洲和亚洲的不断扩展，其在全球市场知名度和影响力方面的执行能力可能会得到提高。

长处

- 增强 (自动化) : H2O 无人驾驶 AI 通过在多个领域提供增强功能来简化 DSML 的采用。它提供了增强的功能工程，包括自然语言处理 (NLP) 功能，可将文本属性转换为功能。同样，模型选择和超参数调整可以自动化。增强用于减少时间序列分析的复杂性或创建最相关统计信息的可视化。可以使用可通过“食谱”目录获得的开源组件来扩展和自定义无人驾驶 AI。
- 可解释性: H2O.ai 通过使用多种技术 (例如 K-LIME, LIME-SUP, Shapley, 可变重要性, 决策树替代, ICE, 部分依赖图, 不同影响分析和“假设”) 提供丰富的可解释性功能树立了榜样分析。” AutoDoc 功能会自动生成文档格式的完整说明集。

- 高性能 ML 组件：H2O.ai 的开源 ML 组件为行业建立了标准，并与许多其他平台集成在一起。商业和开源产品均可在所有主要云上使用。该供应商的组件针对 CPU 多核和多节点配置进行了高度优化和并行化。H2O4GPU 提供了可显著加速 GPU 的软件层。

注意事项

- H2O 无人驾驶 AI 与开源平台之间的凝聚力有限：尽管 H2O.ai 的开源产品线可以通过与 H2O 无人驾驶 AI 的新“配方”功能共享结果，但它们的分离仍然阻碍了技术娴熟的企业用户的潜在协作。在公民数据科学市场中占有较大份额的竞争产品提供了更完整和协作的解决方案。
- 缺少某些数据访问和准备功能：与它在 ML 中强大的功能相比，H2O.ai 在数据访问和数据准备方面有改进的余地。其中包括数据刷新，数据沿袭，访问控制，元数据管理和数据目录，通过 REST API 进行数据访问，加水印，标记和注释。
- 其他高级分析：H2O.ai 似乎没有优先考虑物联网或决策自动化用例所需的补充和替代技术，例如流处理，图形分析，地理空间分析，离散事件或基于代理的仿真，优化实验和决策的设计管理。

IBM 公司

IBM 总部位于美国纽约的 Armonk，其“魔力象限”中考虑的核心产品是 Watson Studio。Watson Studio 的支持产品组合包括 Watson Machine Learning，Watson 知识目录，Watson OpenScale，SPSS（建模和统计），用于 Watson Studio 的 IBM Decision Optimization 和 IBM Streams。

IBM 举行了许多大型用户会议，以及针对特定领域和区域的事件和研讨会。它还参与了多个开源 ML 社区，包括 TensorFlow，PyTorch，Apache Spark 和 Kubeflow。

IBM 使用各种渠道和模型为其全球客户群提供支持。支持计划包括从免费的按需数字支持到人员配备齐全的模式。

IBM 是今年魔力象限中唯一的挑战者。借助 Watson Studio 及其支持产品组合，IBM 证明了随着时间的推移它可以支持客户取得成功。供应商通过 Cloud Pak for Data 展示了混合和多云数据科学领域的思想领导力。改善其产品捆绑和相应的进入市场的方法，将使 IBM 为改进其产品并与来自大型和小型供应商的日益激烈的竞争保持同步而做出的值得称赞的努力。

长处

- 执行 Watson Studio 策略：Watson Studio 履行了产品早期所展示的最初承诺，即所谓的 Data Science Experience。SPSS 的最佳功能（例如，拉票和模型管理）已经很好地集成到了产品中。通过 Watson Studio 产品组合，IBM 在数据管理和信息架构以及灵活性方面保持了牢固的基础。
- 产品部署和业务价值：参考客户对 IBM 的服务和支持表示高度满意。Watson Studio 在可见的企业用例中建立了良好的地位，并为组织带来了价值。IBM 还选择成为最具支持性的混合和多云 DSML 大型供应商之一。IBM 在 Red Hat OpenShift 上用于数据的 Cloud Pak 旨在跨多种部署类型和数据源灵活地管理数据工程和 MLOps。
- 协作并吸引多层次的专业知识：Watson Studio 具有一些新的和经过修改的协作功能，包括目录，资产库，资产沿袭和权限管理。Watson Studio 及其支持的产品组合为不同的分析角色和数据科学专业知识水平提供了强大的用户体验。

注意事项

- 产品捆绑和配置：端到端数据科学所需的 IBM 产品数量正在增长，并且产品组合的开发取决于多个产品团队。客户报告在导航多个产品时遇到的挑战，而产品之间的交换不是无缝的。这些工具的许可成本也仍然是一个问题。
- 在关键领域赶上后对创新的持续需求：IBM 与 Watson Studio 在现代数据科学领域建立了强大的据点。但是，其最近的改进将需要其他创新以保持相关性。例如，在 Watson Studio 中，AutoAI 在增强分析方面迈出了重要的一步，但是许多功能将很快成为“桌面赌注”，其他供应商也将展示创新。
- 品牌恢复和跟上激烈的竞争：Watson Studio 指示了现代数据科学平台的发展方向，但 Watson 尚未成为现代 ML 的顶级品牌。改造 Watson Studio 已经花费了一些时间，IBM 的竞争对手一直很忙。IBM 的挑战是在高级 ML 领域恢复 Watson 品牌的实力，并与传统竞争对手，大型云供应商以及越来越多的小型敏捷供应商保持同步。

尼米

KNIME 总部位于瑞士苏黎世，提供了一个称为 KNIME Analytics Platform 的开源平台。此外，商业扩展 KNIME Server 还提供了包括协作，自动化和操作功能的高级功能。

一个活跃的最终用户社区与 KNIME Hub 完全集成，后者为 KNIME workflow，节点，组件和扩展提供了一个开放的，可搜索的蓝图和最佳实践市场。KNIME 每

年举办两次用户会议，一次在欧洲，一次在美国，以及世界各地的许多聚会和“学习马拉松”。

庞大的全球合作伙伴网络为 KNIME 提供客户支持。也可以通过 KNIME 社区论坛获得支持（英语）。KNIME Server 商业客户具有完整的支持服务合同。

KNIME 被定位为有远见的公司，由于与其他评估的供应商相比，其知名度较低且收入增长缓慢，因此从去年移出“领导者”象限。但是，它继续将重点放在创新，出色的产品和数据科学界内部的深层联系上。

长处

- 开源，为企业级功能提供商业支持：开源 KNIME Analytics Platform 与商业 KNIME Server 相结合，是一个强大，全面且端到端的 DSML 解决方案。开源社区的贡献增强了所提供功能的广度，这使其易于快速扩展。KNIME Server 促进了端到端分析过程中的编排，包括简化的模型操作。
- 支持和支持多种技能水平：KNIME 的传统数据科学功能不仅通过其开放方法得以扩展，而且还具有通过 KNIME Hub 创建并与多种类型的用户共享可重用组件的能力。这些功能可用于以编程为中心的开发人员编码和利用 API，使用经验丰富的数据科学家工作流的专家数据科学模型开发以及增强的公民数据科学家功能。
- 成熟的增强分析愿景：KNIME 的增强分析愿景是明确定义的，可根据用户需求在更具指导性或自动化的方法之间提供灵活性。通过额外的增强功能和对可共享组件的支持，继续专注于数据科学的抽象，从而进一步增强了增强的视野。

注意事项

- 可见性低下且相对增长缓慢：与其他进入者相比，KNIME 在市场上的可见性较低且增长缓慢。尽管 KNIME 开源版本的用户数以千计，但升级到商业平台的人数却很少。因此，许多人无法利用完整产品所提供的端到端功能。在新生和成熟的数据科学团队中，KNIME 仍有很大的发展机会。
- 培训的缺点：尽管 KNIME 既可以提供自我培训，也可以提供面对面的培训，但是培训用户-尤其是缺乏数据科学经验的用户-更具挑战性。此外，入门和培训方面的困难可能会损害整个企业对平台的接受和使用。
- 性能问题：当处理大型数据集，冗长的工作流或复杂使用多个扩展（有时需要大量数据移动或转换）时，性能可能会成为问题。尽管供应商最近已为解决某些性能问题（例如大数据和流节点）做出了一些努力，但客户仍将性能视为关注的领域。

MathWorks

MathWorks 总部位于美国马萨诸塞州纳蒂克。它的两个主要产品是 MATLAB 和 Simulink，但只有 MATLAB 满足此魔力象限的纳入标准（Simulink 被视为支持产品组合的一部分）。

公司的悠久历史和工程科学界的认可，使该公司受益于庞大的用户社区。引用用户赞扬了他们从 MathWorks 获得的支持，并且在实现该技术时常常发现自己是自给自足的。MathWorks 用户的同行社区参与度在魔力象限中是所有供应商中最高的。

MathWorks 被定位为领导者。它从去年在“远见者”象限中的位置转变，是对供应商的适应性以及对尖端技术和长期执行能力的敬意。随着对 ML 和 AI 技术的需求在以资产为中心的领域内不断增长，MathWorks 处于充分利用这一需求的有利条件。随着越来越多的 ML 技术的集成，MathWorks 为渴望快速实现其 ML 工作成果的社区解决了新的业务问题。

长处

- 以生产为重点的环境：MathWorks 与其服务的行业群保持一致，提供了精心设计的环境；从数据预处理和模型开发（通过完全开放的环境）到生产，包括模拟要部署的系统的能力，并可以自动生成代码。然后，可以将代码嵌入到本地的智能资产，边缘或任何交易系统中，云中或通过混合平台。从一开始，公司平台上的每个模型就注定要生产。
- 平台一致性和集成性：MathWorks 提供了一个完全集成的平台，所有新技术（例如强化学习，迁移学习）都可以无缝集成到其环境中。在整个准备，开发，仿真，部署和生产步骤中，MATLAB 提供了完善的体验。
- 预先构建的解决方案：通过其用户在解决特定问题方面的众多实现，MathWorks 打包了诸如预测性维护之类的交钥匙应用程序的经验和领域专业知识。

注意事项

- MLOps：鉴于 MATLAB 具有将模型投入生产的熟练能力，下一步的逻辑步骤是为用户提供监视，管理和治理这些模型的手段。MLOps 功能还需要处理高度分布式的模型。这些是 MathWorks 在即将发布的版本中需要显著增强的功能。
- 数据科学与项目合作中心：笔记本模型的开发对于大多数工程师和数据科学家来说都是必须的。但是，如果 MathWorks 想要扩大其用户范围并改善主题专家或

公民工程师与模型构建者之间的协作, 则它必须提供比笔记本电脑更简单的通信介质, 并提供分析 workflow (画布) 功能。

- 可解释的 AI: 随着 ML 模型的生产遍及其用户, 越来越多的模型使用者 (和监管者) 将开始向 ML 构建者施加压力, 要求他们更好地解释模型。复杂的工程系统特别容易受到这种需求的影响, 而使用深度学习和强化学习技术的 MATLAB 新部署将扩大这种需求。这将要求 MathWorks 大大增强其解释能力。

微软

Microsoft 位于美国华盛顿的雷德蒙德。此魔力象限考虑的核心产品是 Azure 机器学习 (Azure ML)。Azure ML 的产品支持产品组合包括 Azure 机器学习 Studio, Azure 数据工厂, Azure HDInsight, Azure Databricks, Power BI 和其他组件。对于本地工作负载, Microsoft 提供了机器学习服务器。

微软是全球 AI 社区的赞助商, 并且在其全球训练营中取得了显著增长。Azure AI 社区页面为各种背景和技能的用户提供了广泛的论坛和教程。微软还与硬件提供商和 OSS 社区的各种成员保持着长期的合作关系。该供应商全年举办许多大型用户会议, 包括 Microsoft Ignite, Microsoft Inspire 和 Microsoft Build。微软的联合销售, 行业, 部署和咨询合作伙伴数以千计。Microsoft 统一支持为 Azure ML 及其产品组合提供了几种支持选项。

经过一年的努力, Microsoft 的远见卓识和执行能力不断提高, 仍然是一名有远见的人。但是, 它的执行能力受到云优先方法和棘手的一致性问题的限制。微软是本魔力象限中能满足所有数据科学人才的少数供应商之一。

长处

- 具有多种技能水平和分析角色的吸引人和成熟的功能: Microsoft 正在实现其目标, 以广泛的复杂性和要求吸引用户。Azure ML 可以通过拖放和增强分析的形式为市民数据科学家提供大量功能, 而又不会失去其作为专家平台的身份。较大的产品组合提供的产品和功能将吸引开发人员 (较大的 Azure 堆栈), 数据工程师 (Azure Data Factory, Azure Databricks) 和分析师 (Power BI)。
- 发展步伐: 客户欣赏 Microsoft 快速更新和改进云产品的能力。当发布更新时, 用户会感觉到声音, 因此他们的体验会更好。专家用户尤其喜欢始终拥有来自开源社区的新作品和最新作品, 并将其融入他们的工作中。
- 计算能力和控制能力: Microsoft 在性能和可伸缩性方面的得分仍然是所有评估供应商中最高的。Azure ML 具有出色的功能, 可实现成本控制和可见性。许多

客户使用 Azure ML 及其支持产品组合来管理敏捷，大规模和成熟的数据科学操作。

注意事项

- 本地，混合和多云不是一等公民：绝大多数 Azure ML 客户都部署在纯云环境中。在混合，多云或本地环境中，Azure ML 产品组合中的许多功能变得更弱或更复杂。与 Google 和 Amazon Web Services (AWS) 一样，多云支持也很有限。
- 连贯性：尽管 Azure 生态系统为数据科学提供了多种工具和方法，但是许多用户发现大量组件不堪重负，并且对整体体验感到沮丧。混合环境中的数据科学团队面临许可和总体拥有成本的担忧。
- 可靠但不是领先的增强 DSML：Microsoft 继续进行投资，以跟上以数据科学为中心的小型供应商的增强 DSML 功能。增强数据预处理和功能工程等新功能是朝正确方向发展的步骤，但增强 DSML 是该领域发展最快的领域之一。微软对增强分析的长期愿景是正确的，但在短期内，其他主要专注于该领域的供应商正在赢得新客户。

RapidMiner

RapidMiner 位于美国马萨诸塞州的波士顿，RapidMiner Studio 是供应商的主要模型开发工具，既有免费版本，也有商业版本。RapidMiner Server 是企业扩展，旨在部署和维护模型并促进协作。RapidMiner 的支持产品组合还包括 RapidMiner 实时评分和 RapidMiner Radoop。RapidMiner Turbo Prep，RapidMiner Auto Model 和 RapidMiner Automated Model Ops 是该平台的增强功能，而 RapidMiner AI Cloud 提供了基于云的部署选项。

RapidMiner 庞大的用户社区非常活跃，并得到了供应商的支持工作（包括技能认证）的支持。该供应商托管着最大的数据科学社区之一，并且每年在美国和德国举行两次用户会议。

RapidMiner 提供 24/7 企业支持，以根据问题的严重性为付费客户提供 SLA。它还提供了包括白手套服务在内的卓越中心计划。

RapidMiner 被定位为有远见的公司，由于其相对于该魔力象限中其他供应商的增长较慢，因此从去年的领导者象限中撤出。它始终致力于创新和卓越的产品，在竞争激烈的市场中争取市场份额的同时，将为供应商提供良好的服务。RapidMiner 针对企业客户的新使命宣言和新上市策略可能有助于加速增长。

长处

- 端到端增强数据科学生命周期管理: RapidMiner 使专家数据科学家和新兴的公民数据科学家可以轻松地管理从创建到模型构建再到生产的端到端数据科学管道。该供应商通过其 RapidMiner 学院专门提供了一个认证计划, 供非数据科学家了解产品, 模型开发和可操作性。Turbo Prep, Auto Model 和 Model Ops 形成了端到端增强数据科学的引人注目的愿景。
- 灵活性, 敏捷性和现代可扩展性: RapidMiner 强烈支持多种技术 (例如 R, Python, Scala, Java, MATLAB, Octave, HiveQL, Pig, SQL 和 Groovy)。对于深度学习, 将 Keras, TensorFlow, Eclipse Deeplearning4j 和 Theano 集成在一起, 并且 Weka 库中的所有算法都可以作为 RapidMiner 平台的一部分使用。RapidMiner 还允许使用 Docker 和 Kubernetes 进行容器化, 以透明地运行和扩展模型。
- 模型治理: 通过其新的自动化模型操作解决方案, RapidMiner 可以在本地和云中实现对模型的合规性和审计。此治理不仅涵盖数据沿袭, 还涵盖模型训练, 模型使用, 模型再训练, 模型版本控制和模型访问。

注意事项

- 相对增长缓慢和竞争加剧: 与具有类似价值主张的其他供应商相比, RapidMiner 的相对增长缓慢, 尤其是在最近两年中。这部分是由于业务战略优先考虑利润而不是增长的结果。尽管成千上万的用户下载了 RapidMiner, 但升级到商业平台的用户却很少。RapidMiner 仍然是数据科学的基础品牌, 但需要增加收入和市场份额来保持产品开发和相关性。
- 销售创新: RapidMiner 的销售和扩展策略与大多数竞争对手不同, 其收入的增长速度大致与整体市场增长率相同。与其他类似规模的供应商相比, 其销售团队和平均交易规模较小。为了更快地成长, RapidMiner 需要一种新方法将许多尝试该产品的年轻数据科学专业人员转换为付费企业客户。
- 许可和服务定制: 潜在客户发现很难浏览 RapidMiner 的定价模型。总体拥有成本可能难以计算。客户还希望看到更多定制服务。

SAS

SAS 位于美国北卡罗来纳州的卡里市, 它提供用于分析和数据科学的各种软件产品, 支持统计, 机器学习, 文本分析, 预测, 时间序列分析, 计量经济学和优化。SAS 视觉数据挖掘和机器学习 (VDMML) 是此魔力象限评估的核心产品。针对业务分析师和公民/专家数据科学家, VDMML 集成了多种产品, 包括可视化分析和可视化统计。

SAS 为分析提供了专门的资源中心，该资源中心提供了一系列网络研讨会，事件，情况说明书，网络广播，白皮书等。SAS 资源中心既是图书馆又是支持社区，供用户在分析中查找功能信息，提示和技巧以及案例研究。SAS 举办 SAS 全球论坛以及各种客户联系活动。除 SAS Premium 支持和 SAS Elite 支持外，它还提供 24/7 标准技术支持。

SAS 今年再次被定位为领导者。其 DSML 产品具有高度的企业就绪性，并始终为客户提供高商业价值。尽管客户采用 SAS 平台的开源替代方案仍然是竞争威胁，但供应商在市场上仍保持强大的适应能力。SAS 的执行能力继续受到高额许可成本的影响，这导致现有和潜在客户都在探索其他选择。它最近推出了称为 Unified Insights 的生命周期产品捆绑包，以降低许可的复杂性。

长处

- 公认和值得信赖的品牌：SAS 的长期市场占有率和值得信赖的品牌赢得了很多客户的尊重。客户选择 SAS 是因为其企业级平台功能以及对整个分析生命周期（从探索到建模和部署）的支持。
- 模型操作：SAS 提供了最佳的模型操作和管理平台之一。它包括对模型衰减的性能监视，超过阈值时对模型的自动重新训练，通过具有模板和版本控制的集中式模型存储库进行的治理以及 SAS 和开源模型的沿袭。
- 易于使用和增强：SAS VDMML 的拖放界面吸引了希望易于使用的公民数据科学家，同时也支持了喜欢编码的传统数据科学家。该产品提供了有关数据质量和准备的自动化建议，并根据变量分布和其他度量提供可视化建议。

注意事项

- 定价和销售执行：SAS 的定价仍然是那些经常研究成本较低的替代品的客户的关注点。开源工具通常与 SAS 产品一起使用，作为控制成本的一种方式，特别是对于新项目。SAS 引入了更灵活的定价，例如爆裂和无限容量，这对客户更加友好。
- 连贯性：SAS 的完整产品组合仍然很复杂，并且常常使客户感到困惑。一些用例需要许可其他产品（例如用于模型管理的 SAS Model Manager 或用于连接到不同数据源的 SAS / ACCESS）的许可。SAS 最近宣布了生命周期产品捆绑包，以帮助解决许可复杂性问题。
- 营销策略：SAS 在使用 SAS Viya 对其核心平台进行现代化改造方面取得了长足的进步，同时提供了更易于学习/使用的产品，例如 VDMML。但是要赢得传统数据科学家的青睐，供应商将需要应对人们对开源语言和工具缺乏支持的看法。

SAS 还需要增加 SAS Viya 的市场营销，作为行之有效的企业级平台来实施和扩展 ML 计划。

TIBCO 软件

TIBCO 软件总部位于美国加利福尼亚的帕洛阿尔托。近年来，TIBCO 通过收购企业报告和现代 BI 平台供应商（Jaspersoft 和 Spotfire）建立了一个功能全面且功能强大的分析平台（TIBCO Data Science）；描述性和预测性分析平台供应商（Insightful, Statistica 和 Alpine Data）；流分析供应商（StreamBase Systems）；元数据管理功能（乐队网络）；和内存数据平台（SnappyData）。该公司建立了一个多元化的用户社区，该社区聚集在包括 TIBCO Exchange 在内的各种论坛中，用户可以在其中共享和重用分析资产。TIBCO 还通过大量贡献积极参与了许多开源工作。参考客户对 TIBCO 在实施和持续支持方面的支持给予了高度评价。

TIBCO Software 再次整合了其广泛产品组合的功能，再次荣登此魔力象限的领导者。TIBCO 正在简化和精简 TIBCO 数据科学，同时保持平台开放并支持快速移动的 ML 环境。TIBCO 在集成中间件领域的起源使供应商在数据工程和部署方面具有优势，并在许多行业的一系列用例中推动其成功。

长处

- 边缘分析和流传输：TIBCO 继续在物联网中脱颖而出-从使用 TIBCO Data Science 进行边缘集成到 TIBCO Streaming 和开源项目 Flogo，再到在设备上执行深度学习模型。TIBCO 仍然是少数领先于边缘的 ML 生产和执行的 ML 供应商之一，特别是对于以资产为中心的组织。
- 开放性：TIBCO 在其平台的灵活性和开放性方面获得了高分。例如，TIBCO Data Science 可以使用广泛的开源功能集成专有开发的模型，同时在其 workflow 环境中管理这些模型。
- 端到端平台：TIBCO 的端到端体验建立在许多竞争对手缺乏的两个基础上。首先，从频谱的数据管理和准备过程结束，TIBCO 得益于 Spotfire 的 BI 和可视化传统。另一方面，在部署和生产中，TIBCO 可以利用其应用程序集成平台技术（例如 TIBCO BusinessWorks, Project Flogo 和 TIBCO Cloud Mashery）中的专业知识。

注意事项

- MLOps: 尽管在部署和生产方面处于领先地位, 但 TIBCO 在模型化运营管理方面仍有许多工作要做。随着竞争差异化和市场需求转向生产 ML 模型的能力, 目前处于早期版本的 TIBCO MLOps 将必须在下一个版本中赶上领先优势。
- 可解释的 AI : TIBCO 需要继续投资于可解释的 AI。当 ML 模型进入生产应用程序时, 业务用户和工程师需要更高的透明度来解释黑匣子提供的见解。TIBCO 将必须集中其即将到来的开发环境的一部分, 以在可解释的 AI 功能上与市场保持同步。
- 增强型 ML: TIBCO 在采用增强型 ML 功能方面取得了进展, 尤其是通过可视化技术和入门级功能。但是, 随着市场在这一系列功能上的快速发展, 供应商将不得不整合其产品以与市场保持一致。

供应商添加和删除

随着市场的变化, 我们审查并调整了我们对魔力象限的纳入和排除标准。这些调整的结果是, 任何魔力象限中的供应商组合都可能随时间变化。供应商出现在“魔力象限”中的一年而不是第二年并不一定表明我们已经更改了对该供应商的看法。这可能反映了市场的变化, 因此反映了评估标准的变化, 也反映了该供应商的关注点变化。

添加

- 牵牛星
- 掉落
- SAP 已推出了新的 DSML 产品。
- Datawatch (Angoss), 于 2018 年 12 月被 Altair 收购。

纳入和排除标准

我们对该版《魔力象限》的纳入标准进行了一些更改。包含过程包括供应商满足收入阈值和识别参考客户的要求。堆栈排名流程评估了产品对 DSML 典型用例场景的支持程度, 即:

- 业务探索: 这是探索未知并形成假设的经典场景, 需要大量的数据准备, 探索和可视化功能。该方案还可以包括合并增强功能以指导数据准备, 可视化和分析的使用。

- 高级原型制作：此场景描述了何时使用数据科学（尤其是新颖的 ML 技术）显著改善传统的分析方法。传统方法可能是使用人工判断，精确解决方案，历史悠久的启发式方法或遗留数据挖掘方法。高级原型利用以下某些或全部：
 - 更多数据来源
 - 新颖的分析方法（例如深度神经网络，转移学习和强化学习）
 - 大规模计算基础架构
 - 专门的计算机科学和 ML 技能
- 改进生产：在这种情况下，组织已经实施了许多数据科学解决方案并将其交付给业务，但是现在的重点转移到对现有模型进行改进，改进和更新。MLOps 的活动和功能也主要属于生产改进用例，其中平台支持版本的发布激活，监视，性能跟踪，管理，重用，维护和治理。
- 增强型数据科学和机器学习：此用例已从“非传统数据科学”重命名。在这种情况下，使用 ML / AI 来自动化和加速数据科学的关键方面，例如特征工程和模型选择，以及模型可操作性，模型说明，模型调整和管理。这减少了对生成，运行和管理高级分析模型的专业技能的需求。它为公民数据科学家和（应用程序）开发人员打开了 DSML 内容创建的过程。高技能的数据科学家也可以提高生产力，并有更多时间专注于其他任务。

我们使用以下 15 个关键功能在四种用例场景中对供应商的功能进行评分：

- 数据访问：产品在多种类型的数据（如表格，图像，图形，日志，时间序列，音频和文本）中支持数据访问的程度如何？
- 数据准备：产品是否具有大量非编码或编码数据准备功能？
- 数据探索和可视化：该产品是否允许一系列探索性步骤，包括交互式可视化？
- 自动化和扩充：该产品是否促进了特征生成，算法选择，超参数调整和其他关键数据科学任务的自动化？
- 用户界面（UI）：产品是否具有连贯的“外观”和直观的界面，理想情况下，该界面支持可视化流水线组件或可视化构图框架（VCF）？
- 机器学习（ML）：与产品一起易于访问和运送（预包装）的 ML 方法，以及对诸如集成技术（增强，装袋和随机森林）和现代降维方案等现代 ML 方法的支持，范围有多广？
- 其他高级分析：如何将统计，优化，模拟以及文本和图像分析等领域的其他方法集成到开发环境中？

- 灵活性, 可扩展性和开放性: 如何将各种开放源代码库集成到平台中? 用户如何创建自己的功能? 该平台如何与笔记本一起使用?
- 性能和可伸缩性: 如何控制桌面, 服务器和云部署? 如何利用多核和多节点配置?
- 交付: 平台对创建 API 或容器 (例如, 代码, 预测模型标记语言[PMML], 打包应用程序) 的能力的支持程度如何, 这些 API 或容器可用于更快地部署到各种业务场景中?
- 平台和项目管理: 平台提供哪些管理功能 (例如安全性, 计算资源管理, 治理, 项目或实验组织, 审核沿袭和可再现性) ?
- 模型管理: 平台提供了哪些功能来监视和校准数百或数千个模型? 这包括模型测试功能, 例如 K 折交叉验证, 训练验证测试拆分, 曲线下面积 (AUC), 接收器工作特性 (ROC), 损耗矩阵以及并排测试模型 (用于例如, 冠军挑战者[A / B] 测试) 。
- 预设的解决方案: 平台是否提供 “预设的” 解决方案 (例如交叉销售, 社交网络分析, 欺诈检测, 推荐系统, 购买倾向, 故障预测和异常检测), 可以通过图书馆, 市场或画廊?
- 协作: 各种技能的用户如何在同一工作流程和项目上一起工作? 如何将项目归档, 注释并在以后重用?
- 连贯性: 该平台支持整个数据分析管道的直观性, 一致性和集成性如何? 平台本身必须提供元数据和集成功能, 以具备前面的 14 种功能, 并提供无缝的端到端体验, 从而使数据科学家在整个数据分析管道中提高工作效率。这种元功能包括确保尽可能地标准化数据输入/输出格式, 以使组件具有相似且一致的外观, 并确保整个平台上的统一术语。

已审查和修改了与每个关键能力一致的子标准, 以使子能力与适当的整体能力保持一致, 并反映区分解决方案的新发展和关键子能力。

由于 Gartner 方法减少了供应商进行调查和访谈所需的客户推荐数量, 因此降低了客户体验的权重。较低的权重反映出较小的样本量和减少的受访者人数。

如上所述, 我们将 “非传统数据科学” 案例重命名为 “增强数据科学和机器学习”。这与 Gartner 关于增强分析主题的其他研究保持一致, 并解决了对非传统数据科学含义的困惑。

四个用例的权重与去年的魔力象限中的权重相同。

为了符合此魔力象限的资格, 每个供应商必须通过以下评估纳入标准。

纳入标准 1: 数据科学和机器学习平台产品

供应商的 DSML 平台需要:

- 提供构建 DSML 解决方案必不可少的基本功能和高级功能（主要是预测模型和规范模型）
- 支持将这些解决方案整合到业务流程，周围的基础架构，产品和应用程序中
- 在数据和分析管道中为以下所有任务提供各种技能的数据科学家支持：
 - 数据提取
 - 资料准备
 - 数据探索
 - 特征工程
 - 模型创建和培训
 - 模型测试
 - 部署方式
 - 监控方式
 - 保养
 - 合作

供应商还需要能够直接和/或通过商业支持合作伙伴为其 DSML 平台提供技术支持。

纳入标准 2：收入和付费客户数量

评估了三种常见的许可证模型，并针对以下每种正在考虑的核心产品，将每种许可证的收入（和/或客户采用率）进行了合并（如果适用）并根据以下标准进行了评估：

- 永久许可模式：2018 日历年或会计年度的软件许可，维护和升级收入（不包括来自硬件和专业服务的收入）。
- SaaS 订阅模式：2018 日历年或会计年度的年度合同价值 (ACV)，不包括年度合同中包含的任何专业服务。对于多年期合同，仅使用前 12 个月的合同价值进行计算。
- 客户采用率：使用供应商的 DSML 平台（不包括试用版）的活跃付费客户组织的数量。

每个供应商的核心产品都需要具备：

- 在 2018 日历年或财年中，永久许可证收入和 ACV 至少合计 7500 万美元，*或者*

- 至少有 1000 万\$结合在任一 2018 日历年的永久许可证收入和 ACV 或会计年度和当与紧邻的前一日期年度或财政年度相比, 至少 19%的合并收入增长或
- 在 2018 日历年或财政年度以及任何一个财政年度中, 永久许可证收入和 ACV 的总和至少为 500 万美元
- 与上一个日历或会计年度相比, 合并收入至少增长 38%, 或者
- 在 2018 日历或财年获得了 50 个新的公司客户

只有具有通过此纳入标准的核心产品的供应商才被视为第三个纳入标准。

纳入标准 3: 客户人数

接下来, 根据所确定的参考客户对满足纳入标准 1 的要求的供应商进行评估。对于要考虑的每种核心产品, 供应商必须表现出显著的跨行业和跨地域吸引力。计数仅包括使用核心产品的最新版本或 2019 年 8 月之前的 12 个月中发布的版本的活跃唯一客户组织。

跨行业参考客户

每个供应商必须使用生产中考虑的每个 DSML 平台来识别活跃的唯一客户组织。要考虑一个核心产品, 需要 10 个独特的组织。这些必须在生产中具有数据科学解决方案, 并且至少来自以下四个主要行业领域:

- 银行和证券
- 通讯, 媒体和服务
- 教育
- 政府
- 卫生保健
- 保险
- 制造业和自然资源
- 零售
- 运输
- 实用工具
- 批发贸易

跨区域客户数

在每个供应商的参考客户中, 以下三个主要地理区域中的每个必须至少有两个活跃的客户组织:

- 北美

- 欧盟（包括英国）和瑞士
- 世界其他地区

只有通过第 2 个入选标准的供应商才进入第 3 个入选标准。

纳入标准 4：产品能力评分

然后，由 Gartner 分析师评估成功通过前三个标准的供应商，以评估其产品达到 15 种关键功能的程度。

由于可以包含在“魔力象限”中的供应商数量有限，因此只有具有最高 16 至 18 个关键能力得分的供应商产品才能继续进行详细的评估阶段。

如果两个或三个供应商的产品捆绑在一起，我们将它们都包括在内，从而使最大供应商数目达到 18 个。如果三个以上的平台捆绑在一起，我们将使用结合了互联网搜索，Gartner 搜索和 Gartner 客户查询数据的指标以确定哪些供应商的产品具有更大的市场吸引力，并以此为根。魔力象限中绝不会出现超过 18 个供应商。

大约有 70 个供应商被考虑纳入。选择了十六个供应商以最终纳入。

排除标准

其 DSML 平台主要销售给应用程序开发人员或业务分析人员并由其使用的供应商不符合此魔力象限的纳入标准。

荣誉奖

以下列表包括不符合纳入标准或由于缺乏信息而无法验证其纳入资格的著名供应商：

- Amazon Web Services (AWS) ，在 Amazon re: Invent 上宣布，该产品向 Amazon SageMaker (主要面向开发人员和数据科学家) 提供了强大的支持 ML 产品和组件的产品组合（包括用于机器学习的集成开发环境 Amazon SageMaker Studio）。在 2019 年 12 月）。
- Cloudera，它为数据科学团队提供了 Cloudera 机器学习（下一代 Cloudera 数据科学工作台），以加速，管理和扩展 ML 工作流，并为企业 IT 提供保护和管理 ML 工作流的能力。
- FICO 专门从事决策管理，特别适合团队为金融服务用例建立和实施 ML 模型。
- Iguazio 的数据科学平台可帮助企业大规模，实时地开发，部署和管理 AI 应用程序。

- Oracle 提供 SQL, R 和 Python API, 支持 Oracle 机器学习以及广泛的数据和分析产品套件, 以支持数据准备, 可视化, 增强分析, 模型开发和部署以及数据科学生命周期的其他阶段。
- SAP 已对其 DSML 平台 (SAP Data Intelligence) 进行了升级, 重点是企业就绪, 数据管理和治理以及与 SAP 众多数据, 分析和 AI 解决方案的集成。
- Teradata 的, 它提供了 Teradata 的华帝, 有利于企业数据和使用用户的统一视图, 大规模进行预测和分析的规范, 以及自主决策和机器学习。
- 世界编程公司 (World Programming) 提供 WPS Analytics, 这是一个灵活的平台, 它使用针对 Python, R, SQL 和 SAS 的可视化和编程工具来支持分析开发, 治理和部署。

评估标准

执行能力

产品/服务: 在既定市场中竞争和/或服务于既定市场的核心商品和服务。此标准评估当前的产品和服务功能, 质量, 功能集, 技能等。这些可以本地提供, 也可以通过 OEM 协议和合作伙伴关系提供, 如市场定义中所定义以及子标准中所详述。

总体生存能力 (业务部门, 财务, 战略和组织): 此标准包括对组织整体财务状况以及业务部门财务和实际成功的评估。该标准还评估了组织继续提供产品并进行产品投资的可能性, 以及产品在当前投资组合中的地位。

销售执行/定价: 此标准评估组织在所有售前活动中的能力以及支持这些活动的结构。其中包括交易管理, 定价和谈判, 售前支持以及销售渠道的整体有效性。

市场响应能力和跟踪记录: 此标准评估供应商响应机会, 竞争者行动, 客户需求发展以及市场动态变化的响应能力, 改变方向, 保持灵活性并取得竞争成功。它还考虑了供应商对不断变化的市场需求做出响应的历史。

营销执行: 此标准评估旨在传达组织信息以影响市场, 提升品牌, 提高产品知名度并在客户心中建立积极认同的计划的清晰度, 质量, 创造力和功效。这种“思想分享”可以通过宣传, 促销, 思想领导, 社交媒体, 推荐和销售活动的组合来推动。

客户体验: 此标准评估使客户能够通过评估的产品获得预期结果的产品, 服务和/或程序。具体来说, 它考虑了供应商与买方互动的质量, 技术支持和客户支持。除其他外, 还可以评估辅助工具, 客户支持计划, 用户组和 SLA 的可用性。

运营：此标准评估组织实现目标和履行承诺的能力。考虑的因素包括组织结构，技能，经验，程序，系统和其他工具的质量，这些质量使组织能够有效地运作。

表 1：执行评估标准的能力

放大表

•

评估标准	权重
产品或服务	高
总体生存能力	介质
销售执行/定价	低
市场响应度/记录	介质
行销执行	低
客户体验	介质
运作方式	介质

资料来源：Gartner (2020 年 2 月)

愿景的完整性

市场了解：此标准评估供应商了解客户需求并使用该了解来创建产品和服务的能力。拥有清晰的市场愿景，倾听并理解客户需求的供应商可以塑造或增强市场变化。

营销策略：此标准寻找清晰，有区别的消息，这些消息在内部进行持续沟通，并通过社交媒体，广告，客户计划和定位声明进行外部化。

销售策略：此标准寻找使用适当网络（包括直接和间接销售，营销，服务和通信网络）的合理销售策略。它还考虑了可以扩大供应商市场范围，专业知识，技术，服务和客户群的范围和深度的合作伙伴。

提供（产品）策略：此标准寻找一种产品开发和交付方法，该方法着眼于市场差异，功能，方法和功能，以适应当前和未来的需求。

创新：该标准寻找直接，相关，互补和协同的资源，专业知识或资本布局，以进行投资，整合，防御或抢先用途。

表 2：视觉评估标准的完整性

放大表

•

评估标准	权重
市场了解	介质
市场策略	低
销售策略	低
提供（产品）策略	高
商业模式	没有评分
纵向/产业战略	没有评分
革新	高
地理策略	没有评分

资料来源：Gartner（2020 年 2 月）

象限说明

领导者

领导者在 DSML 市场中拥有强大的影响力和重要的思想份额。他们展示了在整个数据探索，模型开发和运营过程中的深度和广度。在提供出色的服务和支持的同时，领导者也要敏捷地应对快速变化的市场状况。使用 Leaders 平台的专家和公民数据科学家的数量在不断增加。

领导者最有能力影响市场的增长和方向。它们适用于大多数行业，地域，数据域和用例，因此对这个市场有扎实的了解和策略。他们不仅可以根椐当前的市场状况专注于有效执行，而且还具有坚实的路线图，可以利用这个快速发展的行业中的新发展和先进技术。他们提供思想领导力和创新差异化，通常会在此过程中破坏市场。

领导者是适合大多数组织评估的合适供应商。但是，他们不应成为唯一经过评估的供应商，因为其他供应商可能会更精确地满足组织的独特需求。领导者提供了高标准的基准，应该与其他人进行比较。

挑战者

挑战者已经建立了存在感，信誉，生存能力和强大的产品能力。但是，他们可能无法表现出与领导者同等的思想领导力和创新能力。

挑战者有两种主要类型：

- 悠久的 DSML 供应商因其稳定性，可预测性和长期的客户关系而成功。这些供应商需要振兴他们的愿景，以与市场发展保持同步，并具有更广泛的影响力和创新能力。如果他们只是继续做自己一直在做的事情，他们的成长和市场地位可能会受到损害。
- 在相邻市场中建立的供应商，例如分析和 BI，数据和分析服务提供商以及开发人员工具市场，这些供应商正在通过扩展其当前平台的解决方案进入 DSML 市场。这些供应商不仅为现有客户而且为新客户提供合理的选择。这些供应商证明他们可以影响这个市场并提供明确的方向和愿景，因此他们有可能成为领导者。但是他们必须避免诱惑，要从表面上快速引入新功能。

挑战者处于成功的位置，可以成功定义当前市场并在当前市场条件下有效运营。但是，他们的远见和路线图可能会因缺乏市场了解，过多地关注短期收益，与策略和产品相关的惯性以及缺乏创新而受到损害。同样，他们的营销努力，地域影响力和知名度可能无法与领导者相提并论。

有远见的人

有远见的人通常是规模较小的供应商或较新的参与者，代表着正在塑造或有可能塑造市场的趋势。但是，可能会担心这些供应商保持有效执行并随其增长而扩展的能力。他们通常在这个市场上并不知名，因此通常没有挑战者和领导者的势头。有远见的人不仅有远见，而且有坚实的支持路线图。他们在满足市场需求方面具有创新性。尽管他们提供的产品通常具有创新性，并且在其提供的功能方面很扎实，但是这些产品的完整性和广度通常存在差距。

有远见的人值得考虑，因为他们可能：

- 代表一个机会来启动创新计划
- 提供一些引人注目的差异化功能，从而提供竞争优势，作为对现有解决方案的补充或替代
- 在产品路线图和方法方面更容易受到影响

但是，有远见的人也对买家构成了潜在的风险更大的选择。在当今竞争激烈的 DSML 市场中，有远见者也可能难以获得动力，发展业务，增加市场份额，实现远见并执行其路线图。它们也可能是收购的目标。

随着有远见的人成熟并证明其执行能力，他们最终可能成为领导者。

利基玩家

利基市场参与者展示出在特定行业或方法中的实力，或与特定技术堆栈完美结合。买方应考虑其特定细分市场。

一些小众玩家展现出一定的远见，这意味着他们可以成为有远见的人。但是，相对于市场上的其他公司，他们通常会努力使自己的愿景引人注目。在驱动和定义市场方面，他们被认为是追随者，而不是领导者。他们可能还难以建立创新和思想领导力的记录，这可能会使他们有动力成为有远见的人。

如果其他利基市场参与者继续以提高其在市场上的势头和吸引力的方式执行任务，则他们可能会成为挑战者。

语境

同时，DSML 市场比以往任何时候都更加活跃和混乱。厂商将迅速发展的专有解决方案与众多开源组件和日益复杂的合作伙伴网络编织在一起。数据和分析领导者力求满足需要尖端功能的专家数据科学家，雄心勃勃的公民数据科学家，寻求快速且可访问的模型构建者以及应对 ML 工程巨大挑战的开发者。这个市场的运动是迅速的和多方向的。快照很快就会过时。

曾经沉睡的 Google 和 AWS 巨头现在在整个领域中醒来并打雷 (尽管 Amazon SageMaker 被排除在此魔力象限之外, 因为它在评估时主要销售给应用程序开发人员并由其使用)。IBM 和 SAS 的传统空间力量正在通过现代产品和自适应策略进入新的十年。许多中小型供应商处于持续高速增长时期。该领域中其他长期受人尊敬的品牌展示了令人兴奋的创新和健康的财务状况。

与去年的“魔力象限”一样, 供应商高度关注创新和差异化, 而不是纯粹的执行力。创新仍然是生存和相关性的关键。即使拥有大量成熟且思想领先的供应商, 仍会涌现出无数具有不同产品和价值主张的新 DSML 初创公司。这个魔力象限中的精选小组已经建立了强大的客户基础, 财务业绩和技术。尽管如此, 新一代供应商仍在努力争取知名度, 相关性, 也许有一天会占据重要的市场份额。考虑到这个市场的竞争和充满活力的性质, 可能已经竞争在象限左侧放置的供应商很难闯入。

在这个市场上的并购活动一直是定期的, 但规模不大。魔力象限中的供应商可能会继续收购有趣的公司来完善其平台, 而转型性的重大收购 (与 Salesforce 在 2019 年收购 Tableau 相同) 始终是可能的。

数据和分析领导者需要努力工作以跟上这个市场。最终用户组织需要增加参与度, 以保持合理的最新状态。领导者应专注于为 DSML 开发新的用例和应用程序-这些用例和应用程序应高度可见, 可以提供真实的业务价值并为未来的计划提供动力。此外, 他们应寻求将市场技术的访问权限扩展到非传统角色, 并制定重要的内部教育计划。

无论最终用户组织是刚刚开始使用预测性和规范性分析, 还是在高级分析中具有成熟的功能, 它们都必须监视市场的变化。这包括供应商如何开发产品并为不同类型的用户和扩展用例提供新功能。企业应首先确定自身产品组合中的差距, 并根据其业务需求监控供应商的产品。他们应该首先评估其现有的分析供应商是否正在应对新的挑战。他们不仅应该考虑 DSML 供应商, 还应该考虑分析和 BI 供应商, 它们正在不断扩展其功能以执行更高级的分析。随着 DSML 功能在企业中越来越多地被采用, 跨部门的工作对于避免过度分散和缺乏通用标准很重要。否则, 各个部门可能会采用不同的平台和流程, 这种情况会导致与运营和维护相关的问题。

为了获得完全成熟的高级分析功能, 组织必须计划并投资端到端数据科学生命周期。生命周期包括访问和转换数据, 进行分析和建立分析模型, 操作和嵌入模型, 随时间推移管理和监视模型以重新评估其相关性以及调整模型以反映数据和业务环境变化的过程。

无论是开始还是扩展 DSML 领域的旅程，组织都不必独自旅行。数据和分析服务提供商提供指导，结构化方法并降低故障风险。服务提供商还帮助缓解数据科学人才招聘和保留方面的常见挑战（请参阅“数据和分析服务提供商的市场指南”）。

市场概况

DSML 市场超越了健康而激动人心的创新。众多供应商提供了广泛的功能，并提供适合大多数成熟度级别的解决方案。数据科学和数据科学家的定义和参数不断发展，其空间与 2014 年“魔力象限”的诞生有很大不同。

现在，许多供应商都希望借助其平台成为最佳选择，以同时吸引专家数据科学家和公民数据科学家，并使他们高兴。随着数据科学生命周期中支持人员的参与越来越普遍，供应商正在添加更多为数据工程师，开发人员和 ML 工程师设计的功能。以前只适合专家数据科学家的供应商正在添加增强的功能和改进的界面，以吸引公民数据科学家。供应商希望扩大其解决方案的覆盖范围和可用性，以最大程度地提高客户的平台投资回报率。

许多供应商将其定位在“视觉完整性”轴的右侧，这表明，仍有大量令人信服的创新和富有远见的路线图。尽管供应商的愿景和价值主张的许多要素重叠，但差异化的关键领域仍在不断涌现。这些包括用户界面，增强的 DSML (autoML)，MLOps，性能和可伸缩性以及尖端的用例和技术（例如，深度学习，大规模 IoT，增强学习）。

许多组织正在使用免费或低成本的开源和公共云服务提供商的产品来启动 DSML 计划，以积累知识并探索可能性。然后，他们很可能会采用商业软件来解决更广泛的用例和团队协作的要求，并实现其模型的部署和管理。尽管使用纯开放源代码堆栈实现企业数据科学的成功是可能的，但绝大多数成熟而有影响力的数据科学团队已经在商业平台上进行了投资。

DSML 平台软件的总收入在 2018 年增长了 19%（高于 2017 年的 17%），代表了分析和 BI 软件市场第二快的细分市场（仅次于现代 BI 平台）。该部门 2018 年的收入为 32 亿美元（高于 2017 年的 26 亿美元）。它在整个分析和 BI 市场中的份额从 2017 年的 14.1% 增长到 2018 年的 15.1%。该领域的一些较小和较年轻的供应商现在正在维持高速增长。与这个魔力象限中的许多供应商相比，以市场速度增长实际上是一个缓慢的增长速度。对这个市场感兴趣的人应该监视并定期评估以下发展：

- DSML 平台的用户组合越来越多样化。专家数据科学家仍然是主要用户，但公民数据科学家已接近代表更大的需求来源和潜在收入。没有标准的公民数据科学家。角色不仅包括业务和 BI 分析师，而且还包括来自传统数据空间的人员，例如数据分析师和数据工程师，以及应用程序开发人员和应用程序工程师。随着越来越多的用户（以不同的角色）采用 DSML 平台，协作和共享的能力比以往任何时候都更为重要。
 - 分析和 BI 平台空间以及 DSML 平台空间都继续碰撞并相互影响。分析和 BI 领域中越来越多的供应商通常通过增强的供应商提供预测和规范功能。就其本身而言，数据科学供应商正在为其平台添加更强大的数据转换和数据可视化功能，同时使其环境对没有传统数据科学背景的个人更加友好。
 - 尽管新的供应商正在进入市场，但“传统”供应商却具有很高的相关性。DSML 领域中的许多传统供应商已经牢固地建立了新产品，或者正在对其方法进行改进和现代化，或者通过战略合作伙伴关系，合并或收购来扩大规模。知名人士继续提供新的功能和方法。同时，它们使现有客户能够继续从他们已经进行的投资以及他们曾经使用过的技术堆栈中受益。
 - 开源生态系统和社区一如既往地充满活力。Python 已牢固地确立为 DSML 中的主导语言，R 社区也在不断发展。开源软件使组织只需很少的前期投资或额外投资即可启动或扩展 DSML 计划。此外，该生态系统向在 DSML 市场中额外提供商业平台的供应商开放并受其支持。
 - 算法构建块通常用于创建模型。随着针对特定领域和行业问题继续对模型进行抽象和打包，这种趋势将继续。
 - 打包的模型越来越多地通过可轻松与应用程序集成和在应用程序中使用的 API 提供（请参阅“Cloud AI 开发人员服务的魔力象限”）。许多云服务 API 高度关注特定领域和行业问题。这种方法可以减少甚至消除组织自身构建模型的需求。
 - 数据科学团队在开发模型方面比在操作模型方面变得更好。结果，往往无法衡量或实现业务价值。数据科学平台中的 MLOps 功能不仅限于部署，还包括生产中模型的持续管理和维护。尽管开发了许多模型，但很少有一种模型能够以不仅导致部署而且导致进行中的管理和维护的方式进行操作。此外，随着业务条件的变化，没有正确管理和监控的模型可能会变得不相关或不准确。
- 在可预见的未来，DSML 市场的变革和创新步伐将继续加快。
- 在过去的一年中，DSML 平台产品的几种趋势得到了发展。现代平台包含或容纳以下内容：

- 组件化：当供应商开发自己的组件，使用开源软件或与其他供应商合作以扩展其产品时，由多个组件组成的平台已成为常态。与单个产品中的本机集成相反，供应商越来越多地提供了各种各样的工具集合。DSML 平台的定义已在今年进行了重大更新，以反映这一现实。
- 开源验收：尽管程度不同，所有 DSML 平台都使用并结合了开源软件。一些提供用于访问常见开源库的 API。一些公司将开放源代码技术构建为可在其自己的平台中访问的功能。其他功能包括使用开源生态系统中平台内创建的分析工件的能力。还有一些工具提供了更多的包装器，以便在一致的环境中与开放源代码工具进行本机工作，这也可以实现操作性。通过各种协作和精心设计的方法来支持开源平台和框架已成为标准。这些自适应平台增加了对新功能的支持并增加了工作量，同时减少了用户为不同环境切换平台的需求。
- 多种用户类型：数据科学的有机发展要求工具能够满足具有不同技能和不同 DSML 知识水平的用户的需求。某些平台仍主要是针对特定用户类型（通常是专家数据科学家）设计的。但是，使范围广泛的用户（从公民数据科学家到专家数据科学家再到应用程序开发人员）的组件或功能正在越来越普遍。
- 平台一致性：组件化程度的提高和开放源代码的合并增加了零散，笨拙的解决方案的潜力。为了获得完整，强大的功能而需要访问多个组件和平台的需求，必须与通过无缝且一致的方式访问所有功能的需求相平衡。随着产品拥抱异构环境，凝聚力变得越来越重要。随着产品的扩展以提供更多功能并与新兴技术保持同步，至关重要，它们不仅要支持管理多个组件的能力，而且还需要从平台内部轻松，无缝地访问它们的能力。
- MLOps：运营能力不仅可以部署，还可以随着时间的推移管理和维护模型。随着业务需求，优先级和条件的变化，MLOps 功能对于鼓励随着时间的推移不断重新评估分析的相关性和有效性至关重要。随着 DSML 脱离实验室并进入主流，必须通过无缝集成以及经过精心设计的体系结构和流程进行操作。MLOps 功能还应包括可解释性，模型版本控制和业务影响分析等。
- 模型和数据存储库：趋势是提供一种跟踪和共享数据以及模型开发和部署过程中生成的分析工件的方法。这对于数据科学计划的工作，治理和企业可伸缩性的重复数据删除至关重要。它还支持正在使用的分析资产的持续更新，并为数据科学操作提供至关重要的透明度。
- 协作：随着对 DSML 平台的访问变得民主化，并且越来越多的用户在整个分析流程中一起工作，能够轻松无缝地进行协作的需求显著增加。随着新型用户对平台的使用越来越容易，这些产品必须使人们能够一起工作并在整个数据科学生命

周期中进行实时共享。DSML 平台还促进了数据科学团队与 IT 之间以及数据科学家与业务线负责人之间至关重要的协作。

- 扩展到决策管理：DSML 平台越来越超出可操作性来支持协作，随着分析工具超越了预测来明确地驱动业务决策，这又激发了人们对决策管理功能的兴趣。

证据

Gartner 在此“魔力象限”中的评估和评论来自以下来源：

- 所选供应商的说明手册和文档。我们使用这些来验证平台功能。
- 从 2019 年 9 月到 2019 年 10 月对供应商的参考客户进行了在线调查。该调查针对参考客户在供应商的平台上的使用情况进行了 175 次答复。从供应商提供的信息中得出的调查参与者列表。
- 使用供应商提供的客户参考进行电话采访。
- 供应商填写的调查表。
- 有关各个供应商的策略和运营的供应商简介，包括产品演示。
- 广泛的 RFP，询问每个供应商如何提供与我们的 15 个关键功能相对应的特定功能（请参阅“工具包：数据科学和机器学习平台的 RFP”）。
- 准备的视频演示了供应商的 DSML 平台如何在 15 个关键功能中满足特定功能要求。
- Gartner 分析师与 Gartner 客户之间的互动决定了他们的评估标准，以及 Gartner 客户对供应商如何成功满足这些标准的看法。

注 1 开源平台的定义

在整个 DSML 平台市场中，开源方法正变得越来越普遍。它使人们能够进行协作创新，从而以缩短上市时间的方式贡献自己的观点。

开源方法正迅速成为引入新功能的主流方法。此魔力象限中评估了许多此类功能。DSML 平台市场中最常见的开源示例是组件。

开源组件包括：

- 开源编程语言，例如 Python 和 R
- 开源库和框架，例如 scikit-learn 和 TensorFlow
- 开源可视化文件，例如 D3 和 Plotly
- 开源笔记本，例如 Jupyter 和 Zeppelin
- 开源数据管理平台，例如 Apache Spark 和 Hadoop

如果平台提供了访问开源组件的灵活性和可扩展性，则该平台被认为是开放的，但不是开源的。此外，平台本身可以是开源的，这意味着可以使用或修改其源代码。

开源软件通常是作为公共协作开发的，可以免费获得。但是，只有具有商业许可产品的开源平台才有资格包含在此“魔力象限”中。

评估标准定义

执行能力

产品/服务：供应商为定义的市场提供的核心商品和服务。这包括当前的产品/服务功能，质量，功能集，技能等，无论是本地提供还是通过市场定义中定义的和子标准中详细说明了 OEM 协议/合作伙伴关系提供。

总体生存能力：生存能力包括对整个组织财务状况，业务部门财务和实际成功的评估，以及各个业务部门将继续投资该产品，继续提供该产品并改善其状态的可能性。组织产品组合中的艺术。

销售执行/定价：供应商在所有售前活动中的功能以及支持这些功能的结构。这包括交易管理，定价和谈判，售前支持以及销售渠道的整体有效性。

市场响应能力/记录：随着机会的发展，竞争对手的行动，客户需求的发展以及市场动态的变化，能够做出响应，改变方向，保持灵活性并取得竞争性成功。此标准还考虑了供应商的响应历史。

市场营销执行：计划的清晰度，质量，创造力和功效，旨在传达组织的信息以影响市场，促进品牌和业务，提高产品知名度并在产品/品牌和组织中建立积极的

认同感。买家的想法。这种“思想分享”可以通过宣传, 促销计划, 思想领导力, 口口相传和销售活动来实现。

客户体验: 关系, 产品和服务/计划, 使客户能够成功评估产品。具体来说, 这包括客户获得技术支持或客户支持的方式。这也可以包括辅助工具, 客户支持计划 (及其质量), 用户组的可用性, 服务级别协议等。

运营: 组织实现其目标和承诺的能力。因素包括组织结构的质量, 包括技能, 经验, 程序, 系统和其他使组织能够持续有效地运作的工具。

愿景的完整性

市场了解: 供应商了解买方需求和将其转化为产品和服务的能力。表现出最高远见的供应商会倾听并了解买家的需求, 并可以通过增加远见来塑造或增强他们的需求。

营销策略: 清晰, 差异化的消息集, 在整个组织中始终如一地传达, 并通过网站, 广告, 客户计划和定位声明进行外部化。

销售策略: 使用直接和间接销售, 市场营销, 服务和沟通分支机构的适当网络的产品销售策略, 可扩展市场范围, 技能, 专业知识, 技术, 服务和客户群的范围和深度。

提供 (产品) 策略: 供应商的产品开发和交付方法强调差异, 功能, 方法和功能集, 以适应当前和未来的需求。

业务模型: 供应商基本业务命题的合理性和逻辑性。

垂直/行业策略: 供应商的策略, 用于引导资源, 技能和产品以满足包括垂直市场在内的各个细分市场的特定需求。

创新: 用于投资, 合并, 防御性或先发制人的资源, 专业知识或资本的直接, 相关, 互补和协同的布局。

地理策略: 供应商的策略, 可以直接或通过适合于该地理和市场的合作伙伴, 渠道和子公司来引导资源, 技能和产品以满足“本国”或本地地理之外的特定地理需求。