



使用 Amazon A SageMaker I 预测货运能力需求

AWS 规范性指导



AWS 规范性指导: 使用 Amazon A SageMaker I 预测货运能力需求

Copyright © 2026 Amazon Web Services, Inc. and/or its affiliates. All rights reserved.

Amazon 的商标和商业外观不得用于任何非 Amazon 的商品或服务，也不得以任何可能引起客户混淆、贬低或诋毁 Amazon 的方式使用。所有非 Amazon 拥有的其他商标均为各自所有者的财产，这些所有者可能附属于 Amazon、与 Amazon 有关联或由 Amazon 赞助，也可能不是如此。

Table of Contents

简介 1

架构 3

数据 4

内部数据 4

外部数据 4

机器学习模型 5

输入要素预测的 ML 模型 5

目标变量预测的 ML 模型 5

用户输入 6

最佳实践 7

后续步骤 8

资源 9

Amazon Quick 文档 9

亚马逊 SageMaker AI 文档 9

AWS 代码示例 9

文档历史记录 10

..... xi

使用 Amazon A SageMaker I 预测货运能力需求

贾天霞和陈恒志，Amazon Web Services (AWS)

2024 年 5 月 ([文档历史记录](#))

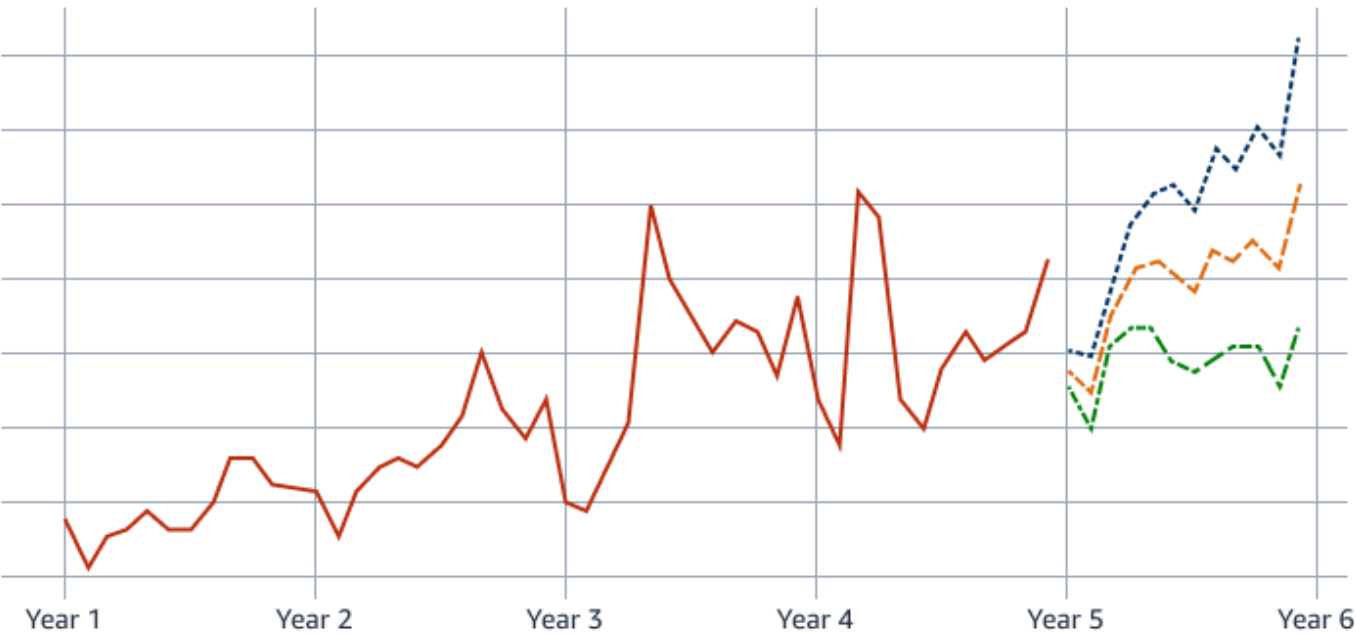
需求预测对于运输和物流行业至关重要，尤其是在供应链紧张时期。准确的货运需求估算使参与物流和供应链的公司受益，例如那些跨境运输集装箱和航空货物的公司。这可以帮助公司有效地管理保护其运输网络的成本，从而帮助他们管理运输成本并最大限度地提高收入和利润。

能够做出准确预测的机器学习 (ML) 模型取决于高质量的训练数据。对于需求预测，培训数据可以包括历史需求量和与其他内部生成的可能与数量相关的数据，例如价格、库存和销售团队人数。此外，外部数据，例如竞争对手、市场环境、假期、天气和宏观经济，也可能影响需求量。这些内部和外部数据因子可用作 ML 模型中的特征。

确定所有功能后，企业可能还想为其控制范围内的某些功能提供意见。例如，企业可以提前设定运费或决定何时进行促销和折扣。在进行预测时，可以将这些类型的用户输入纳入模型中。

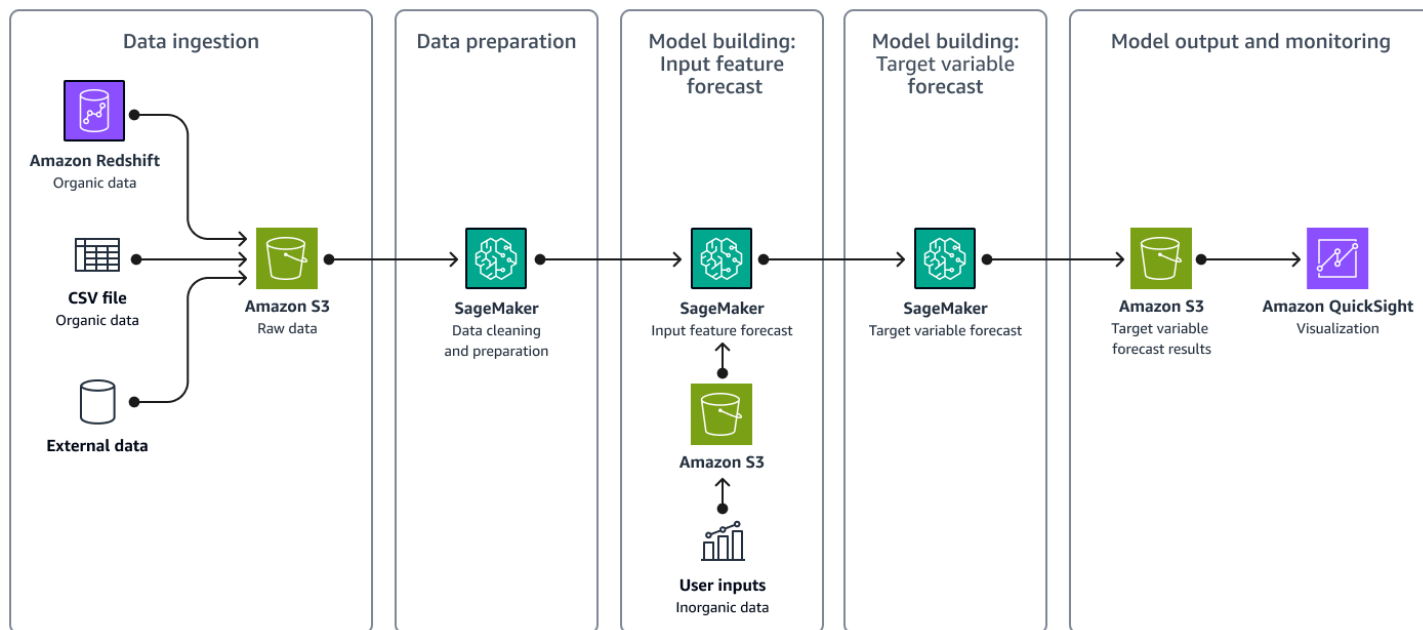
本指南描述了一种构建解决方案的策略 AWS，该解决方案使用机器学习模型进行准确的物流需求预测。您可以在包含需求量和与需求相关的特征的历史数据集上训练模型。这些功能和指标包括内部自然数据和外部数据。该解决方案还允许用户和业务分析师灵活地提供输入，然后将其整合到预测模型中。

下图显示了历史时间序列和 12 个月预测范围的示例。您可以使用本指南中的建议来创建生成此类预测的 ML 模型。



预测货运需求的架构

下图显示了解决方案的工作流程，包括数据摄取、数据准备、模型构建以及最终输出和监控。



该解决方案架构包括以下主要组件：

1. 数据摄取 — 您将自然数据和外部数据存储于亚马逊[简单存储服务 \(Amazon S3\) Service](#) 中。
2. 数据准备 — [Amazon SageMaker AI](#) 会清理数据并为机器学习模型训练做好准备。有关更多信息，请参阅 SageMaker AI 文档中的[准备数据](#)。
3. 模型构建：输入要素预测 — SageMaker AI 用于[Prophet](#)为每个输入要素生成时间序列预测。您可以检查预测结果。如果需要，您可以提供用户输入以覆盖该功能的预测。
4. 模型构建：目标变量预测 — SageMaker AI 使用修改后的输入特征创建用于推理的回归模型。
5. 模型输出和监控-回归模型将预测结果输出到 Amazon S3。您可以在 [Amazon Quick](#) 中对预测进行可视化。分析师可以通过将预测与实际需求量进行比较来监控预测结果并评估准确性。

从数据摄取到最终模型输出的整个处理管道可以编排为自动运行。例如，您可以将其设置为每月自动运行以进行每月需求预测。如果您需要对多个产品进行预测，则可以并行运行多个产品的渠道。有关更多信息，请参阅 SageMaker AI 文档中的[实现 MLOP](#)。

用于预测货运需求的数据

对于任何机器学习模型进行有意义的预测和预测，高质量的数据都是必不可少的。对于需求预测，该数据集包含任何可能影响最终需求的相关数据。这些数据可能来自各种来源。您可以将这些数据分为两类，即内部数据和外部数据。

内部数据

内部数据是业务生成的有机数据。这些数据通常存储在数据仓库中，例如 [Amazon Redshift](#)。

您可以直接从数据仓库中包含感兴趣产品的历史交易量的表中生成或提取目标输出值。对于航运公司，产出或目标值可以是海运的整箱装载量或航空货物的总重量单位。

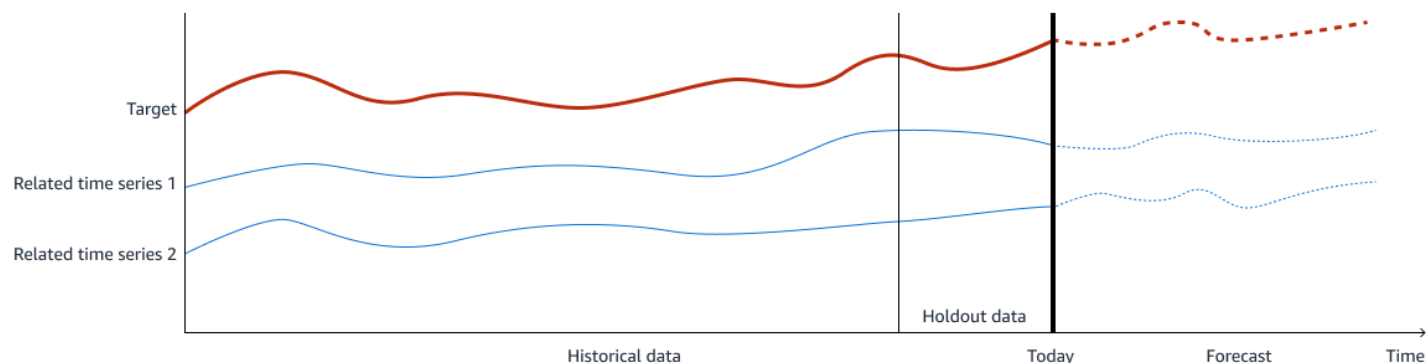
您还可以生成各种历史业务指标。在预测需求时，这些可用作机器学习模型中的功能。示例功能包括历史价格、成本、容量和库存。

外部数据

外部数据源可用作附加功能，以提高预测的准确性。外部数据源的示例包括天气数据、宏观经济数据、行业数据和市场数据。这些因素可能对物流和运输行业产生直接或间接影响，从而影响需求。例如，市场运费为全球货运市场提供了基准，这最终会影响公司的特定需求。宏观经济数据，例如主要经济体的进出口数据，也可以用作衡量市场活动的指标。要整合这些外部数据源，您可以使用各种 API 来摄取数据。例如，St. Louis Fed 提供了 [Federal Reserve Economic Data \(FRED\) API](#) 访问宏观经济数据的权限，National Oceanic and Atmospheric Administration (NOAA) 提供了 [Climate Data Online \(CDO\) API](#) 访问全球天气数据的权限。

用于预测货运需求的机器学习模型

下图显示了训练数据的示例。目标就是你想要预测的内容，而相关的时间序列 1 和 2 是与预测目标相关的输入要素。历史数据用于训练和验证，您保留一段时间的历史数据用于模型验证。



在需求预测中，产出（或目标）是您要预测的需求量。输入要素是与输出相关的时间序列数据。要训练机器学习模型以准确预测需求量，解决方案中需要两个机器学习模型。第一个模型对输入要素进行时间序列预测，包括内部和外部数据。第二种模型使用所有功能进行最终需求预测。通过将这两个模型结合使用，您可以有效地捕捉时间序列趋势以及目标与输入之间的关系。

输入要素预测的 ML 模型

输入要素包括内部和外部历史时间序列数据。要对每个特征进行预测，可以使用一维 (1D) 时间序列模型。有多种算法可供选择。例如，对于具有强烈季节效应的时间序列和多个季节的历史数据，[Prophet](#) 效果最好。系统会为每个单独的功能生成预测。

目标变量预测的 ML 模型

输出或需求量的机器学习模型旨在捕捉所有功能与输出之间的关系。您可以使用各种监督回归模型，例如 lasso、ridge regression、random forest、和 XGBoost。在构建模型并寻找最佳参数和超参数时，可以使用保留数据。Holdout 数据是带有标签的历史数据的一部分，这些数据从用于训练机器学习模型的数据集中保留。通过将预测值与抵制数据进行比较，您可以使用抵制数据来评估模型性能。

用于预测货运需求的用户输入

尽管大多数功能的未来尚不清楚，但可能有一些功能是企业可以控制的。例如，价格是一个通常与需求量有密切关系的特征。由于企业设定价格，因此您知道价格何时会上涨或何时会有折扣或促销。此外，销售团队的规模也会影响需求量，企业可以控制其销售团队的规模。对于企业管理的这些数据点，您可以提供用户输入。在机器学习建模步骤中，一维时间序列模型给出每个要素的预测，然后用户可以检查这些预测值并用用户输入将其覆盖。然后在模型中进行最终预测时使用这些被覆盖的输入。

在一维时间序列模型预测与企业对功能未来行为的预期不匹配的情况下，此用户输入步骤可能至关重要。您可以改写这些预测值，这样可以改善总体产出预测。

预测货运需求的机器学习模型的最佳实践

通过遵循这些最佳实践，您可以提高用于预测货运需求的机器学习模型的准确性、可靠性和可解释性，从而最终实现更好的决策和运营效率：

- **数据质量和预处理**-确保用于训练模型的数据质量高，没有错误、缺失值和不一致之处。数据预处理步骤，例如处理缺失值、异常值检测和特征工程，在提高模型的准确性方面起着至关重要的作用。
- **足够的历史数据**-拥有足够的历史数据对于捕捉模式、趋势和季节性至关重要。但是，考虑历史数据的相关性和及时性也很重要。如果市场、业务运营或外部因素发生了重大变化，则较旧的数据可能无法代表当前情况。在这种情况下，请对最新数据给予更高的权重。
- **特征选择和工程** — 识别相关特征并根据现有数据设计新特征可以显著提高模型的性能。与领域专家密切合作，在选择合适的功能时运用他们的知识和见解。此外，可以考虑进行特征重要性分析，以确定最具影响力的特征，并有可能移除多余或不相关的特征。
- **集成模型** — 与其依赖单个模型，不如考虑使用组合多个模型预测的集成技术。Ensemble 模型的性能可以优于单个模型，并且可以提供更强大、更准确的预测。
- **模型评估和验证**-使用适当的指标，例如均方误差 (MSE)、平均绝对百分比误差 (MAPE) 或任何其他特定领域的指标，定期评估和验证模型的性能。使用交叉验证或保留验证来评估模型的概化能力。
- **持续监控和再培训** — 由于经济状况、市场动态或业务运营变化等各种因素，货运需求模式可能会随着时间的推移而发生变化。持续监控模型的性能，并定期使用最新数据对其进行重新训练，以提高其准确性和相关性。
- **可解释的人工智能** — 需求预测模型应该是可解释和可解释的，尤其是在利益相关者需要了解预测背后的原因的情况下。特征重要性分析、部分依赖图、Shapley Additive 解释 (SHAP) 等技术可以帮助解释模型的决策。
- **整合领域知识** — 与领域专家和业务利益相关者密切合作，将他们的知识和见解融入建模过程。他们的领域专业知识可以帮助识别潜在的偏见，解释结果，并根据预测做出明智的决策。
- **情景分析和假设模拟** — 将执行情景分析和假设模拟的功能整合到预测解决方案中。这使利益相关者能够探索不同的业务决策或外部因素对需求预测的影响，从而做出更明智的决策。
- **自动化且可扩展的管道** — 为数据摄取、预处理、模型训练和部署构建自动化、可扩展的管道。这可以始终如一地高效地执行预测流程，尤其是在处理多个产品或地区时。

后续步骤

在实施此需求预测解决方案之前 AWS，建议您评估要解决的问题。将企业主和数据科学家聚集在一起集思广益，讨论机器学习模型能否解决问题是个好主意。了解您拥有哪些数据集以及可用的历史数据长度至关重要。对于企业主来说，与数据科学家合作以提供领域知识、识别有用的功能并帮助创建这些功能也很重要。模型的可靠性会随着您可以创建的相关特征数量的增加而提高，从而提供更准确的预测。

要在此基础上构建此架构 AWS，首先要设置 AWS 账户 并配置必要的服务，例如用于数据存储的亚马逊简单存储服务 (Amazon S3) 和用于机器学习模型训练的 Amazon SageMaker。接下来，识别并收集将用作预测模型输入要素的内部和外部数据源。将这些数据存储在 Amazon S3 中，并使用 SageMaker 中的数据处理功能为模型训练预处理和准备数据。在 SageMaker 中，使用自动模型调整和分布式训练功能来训练和优化预测模型。您也可以使用 AWS 服务 例如 AWS Step Functions 或 AWS Lambda 来设置一条管道，定期使用最新数据重新训练预测模型。重新训练后，在 SageMaker 中启动批量转换任务以生成预测结果，然后将其存储在 Amazon S3 中。使用 Amazon QuickSight 可视化和监控批量转换作业生成的预测结果。

资源

Amazon Quick 文档

- [什么是 Amazon Quick ?](#)
- [使用数据源](#)

亚马逊 SageMaker AI 文档

- [什么是亚马逊 A SageMaker I ?](#)
- [准备数据](#)
- [执行自动模型调整](#)
- [使用批量转换](#)

AWS 代码示例

- [亚马逊 SageMaker AI 示例存储库](#) (GitHub)

文档历史记录

下表介绍了本指南的一些重要更改。如果您想收到有关 future 更新的通知，可以订阅 [RSSFeed](#)。

变更	说明	日期
初次发布	—	2024 年 5 月 14 日

本文属于机器翻译版本。若本译文内容与英语原文存在差异，则一律以英文原文为准。