



使用 Amazon SageMaker AI 預測貨運容量的需求

AWS 方案指引



AWS 方案指引: 使用 Amazon SageMaker AI 預測貨運容量的需求

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

Amazon 的商標和商業外觀不得用於任何非 Amazon 的產品或服務，也不能以任何可能造成客戶混淆、任何貶低或使 Amazon 名譽受損的方式使用 Amazon 的商標和商業外觀。所有其他非 Amazon 擁有的商標均為其各自擁有者的財產，這些擁有者可能附屬於 Amazon，或與 Amazon 有合作關係，亦或受到 Amazon 贊助。

Table of Contents

簡介	1
Architecture	3
資料	4
內部數據	4
外部資料	4
機器學習模型	5
輸入功能預測的 ML 模型	5
目標變數預測的 ML 模型	5
使用者輸入	6
最佳實務	7
後續步驟	8
Resources	9
Amazon Quick 文件	9
Amazon SageMaker AI 文件	9
AWS 程式碼範例	9
文件歷史紀錄	10
.....	xi

使用 Amazon SageMaker AI 預測貨運容量的需求

Tianxia Jia 和 Hengzhi Chen , Amazon Web Services (AWS)

2024 年 5 月 ([文件歷史記錄](#))

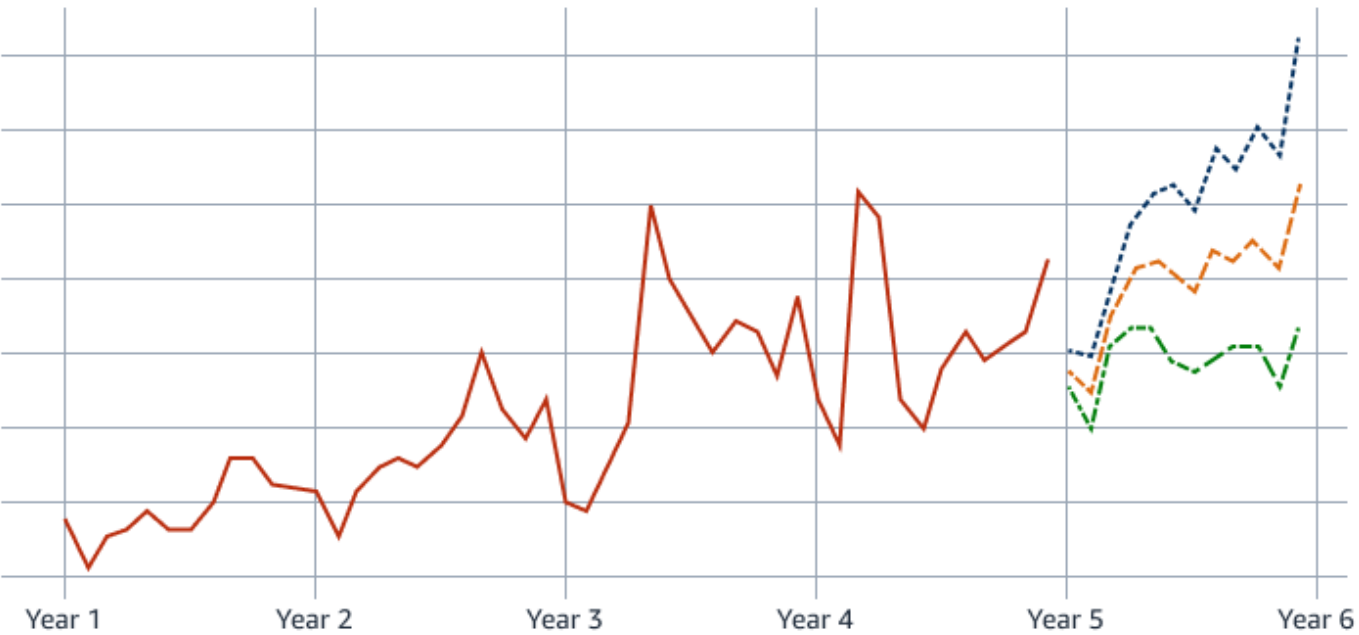
需求預測對運輸和物流產業至關重要，尤其是在供應鏈限制期間。準確的貨運需求預估會讓涉及物流和供應鏈的公司受益，例如跨越邊界運送容器和航空貨運的公司。這有助於公司有效地管理保護其運輸網路的成本，這有助於他們管理運輸成本並最大限度地提高收入和利潤。

能夠進行準確預測的機器學習 (ML) 模型取決於高品質的訓練資料。對於需求預測，訓練資料可以包含歷史需求量和與其他與量相關的內部產生的資料，例如價格、庫存和銷售團隊人數。此外，外部資料，例如競爭對手、市場環境、假日、天氣和巨集經濟，也可能會影響需求量。這些內部和外部資料因素可以用作 ML 模型中的功能。

識別所有功能後，企業也可能想要提供部分這些功能在其控制範圍內的輸入。例如，企業可以預先設定運送價格，或決定何時執行促銷和折扣。這些類型的使用者輸入可在進行預測時併入模型。

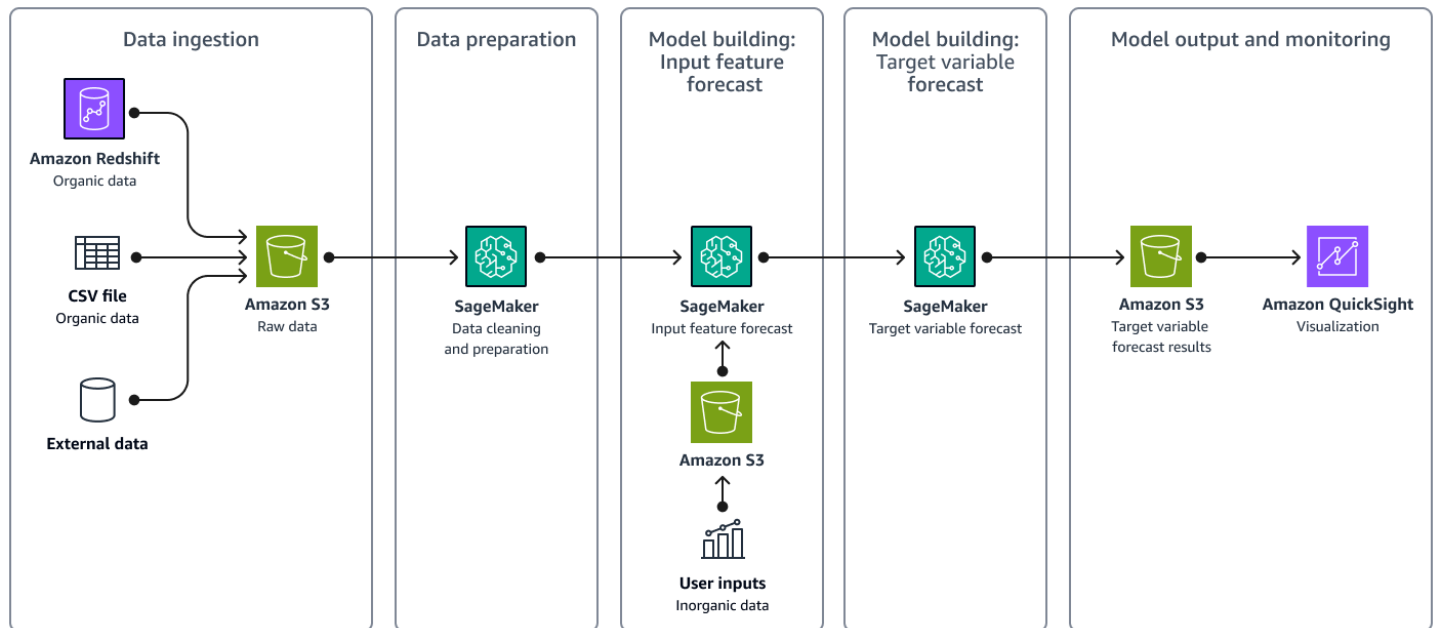
本指南說明在 上建置解決方案的策略 AWS ，該解決方案使用 ML 模型進行準確的邏輯需求預測。您可以在歷史資料集上訓練模型，其中包含需求量和與需求相關的功能。這些功能和指標包括內部有機資料和外部資料。解決方案也為使用者和業務分析師提供了提供輸入的靈活性，然後可以將其納入預測模型中。

下圖顯示歷史時間序列和 12 個月預測範圍的範例。您可以使用本指南中的建議來建立產生此類預測的 ML 模型。



預測貨運需求的架構

下圖顯示解決方案的工作流程，包括資料擷取、資料準備、模型建置，以及最終輸出和監控。



解決方案架構包含下列主要元件：

1. 資料擷取 – 您可以將有機資料和外部資料存放在 [Amazon Simple Storage Service \(Amazon S3\)](#) 中。
2. 資料準備 – [Amazon SageMaker AI](#) 會清除資料並準備 ML 模型訓練。如需詳細資訊，請參閱 SageMaker AI 文件中的 [準備資料](#)。
3. 模型建置：輸入功能預測 – SageMaker AI 使用 [Prophet](#) 為每個輸入功能產生時間序列預測。您可以檢查預測結果。如有需要，您可以提供使用者輸入來覆寫功能的預測。
4. 模型建置：目標變數預測 – SageMaker AI 使用修改的輸入功能建立推論的迴歸模型。
5. 模型輸出和監控 – 迴歸模型會將預測結果輸出至 Amazon S3。您可以在 [Amazon QuickSight](#) 中視覺化預測。分析人員可以透過比較預測與實際需求量來監控預測結果並評估準確性。

從資料擷取到最終模型輸出的整個處理管道都可以協調以自動執行。例如，您可以將其設定為每月自動執行每月需求預測。如果您需要對多個產品進行預測，您可以平行執行多個產品的管道。如需詳細資訊，請參閱 SageMaker AI 文件中的 [實作 MLOps](#)。

用於預測運費需求的資料

高品質的資料對於任何機器學習模型進行有意義的預測和預測都至關重要。對於需求預測，資料集包含任何可能影響最終需求的相關資料。這些數據可以來自各種來源。您可以將此資料分類為內部和外部資料兩類。

內部數據

內部資料是業務產生的有機資料。這些資料通常會儲存在資料倉儲中，例如 [Amazon Redshift](#)。

您可以從包含感興趣產品的歷史磁碟區的資料倉儲中，直接從資料倉儲中產生或擷取目標輸出值。對於航運公司而言，輸出或目標值可以是以海運的完整集裝箱負載為單位，也可以是空運貨物的總重量。

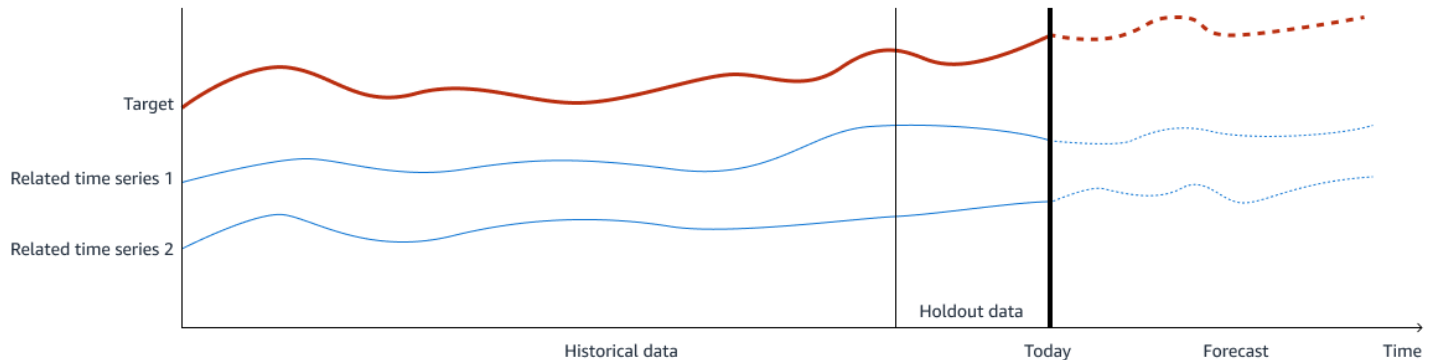
您也可以產生各種歷史業務指標。預測需求時，這些功能可用作機器學習模型中的功能。範例功能包括歷史價格、成本、產能和庫存。

外部資料

外部數據源可用作附加功能，以提高預測準確性。外部數據源的例子包括天氣數據，宏觀經濟數據，行業數據和市場數據。這些因素可能對物流和運輸業產生直接或間接的影響，因此影響了需求。例如，市場運費率提供了全球貨運市場的基準，最終會影響公司的特定需求。宏觀經濟數據（例如主要經濟體的進出口數據）亦可用作衡量市場活動的指標。若要合併這些外部資料來源，您可以使用各種 API 來擷取資料。例如，St. Louis Fed提供了訪問宏觀經濟數據，並National Oceanic and Atmospheric Administration (NOAA)提供了訪問全球天氣數據。[Federal Reserve Economic Data \(FRED\)](#) [APIClimate Data Online \(CDO\) API](#)

用於預測貨運需求的機器學習模型

下圖顯示訓練資料的範例。目標是您想要預測的內容，而相關的時間序列 1 和 2 是與預測目標相關的輸入功能。歷史資料用於訓練和驗證，而您保留一段時間的歷史資料以進行模型驗證。



在需求預測中，輸出（或目標）是您想要預測的需求量。輸入功能是與輸出相關的時間序列資料。若要訓練 ML 模型以準確預測需求量，解決方案中需要兩個機器學習模型。第一個模型會針對輸入功能進行時間序列預測，包括內部和外部資料。第二個模型使用所有功能進行最終需求預測。透過同時使用這兩個模型，您可以有效擷取時間序列趨勢以及目標與輸入之間的關係。

輸入功能預測的 ML 模型

輸入功能包括內部和外部歷史時間序列資料。若要預測每個功能，您可以使用一維 (1D) 時間序列模型。有各種演算法可用。例如，[Prophet](#) 最適合使用具有強烈季節性效果和數個季節歷史資料的時間序列。系統會為每個個別功能產生預測。

目標變數預測的 ML 模型

輸出的 ML 模型或需求量，是為了擷取所有功能與輸出之間的關係而建置。您可以使用各種監督式迴歸模型，例如 lasso、random forest、ridge regression 和 XGBoost。建置模型並尋找最佳參數和超參數時，您可以使用保留資料。保留資料是歷史標籤資料的一部分，從用於訓練機器學習模型的資料集保留。您可以使用保留資料，透過比較預測與保留資料來評估模型效能。

用戶輸入用於預測貨運需求

儘管大多數功能的期貨尚未知，但可能有一些業務可以控制的功能。例如，價格是一個功能，通常與需求量有很強的關係。由於業務設定價格，因此您知道價格何時會上漲，或何時會有 discount 或促銷。此外，銷售團隊的規模也會影響需求量，並且企業控制其銷售團隊的規模。對於企業管理的這些資料點，您可以提供使用者輸入。在 ML 建模步驟中，1D 時間序列模型會提供每個功能的預測，然後使用者可以檢查這些預測值，並使用使用者輸入覆寫這些預測值。然後在進行最終預測時，會在模型中使用這些已覆寫的輸入。

在 1D 時間序列模型預測不符合企業對功能 future 行為的期望時，此使用者輸入步驟可能很重要。您可以覆寫這些預測值，以改善整體輸出預測。

預測貨運需求的 ML 模型最佳實務

透過遵循這些最佳實務，您可以增強機器學習模型的準確性、可靠性和可解釋性，以預測貨運需求，最終實現更好的決策和營運效率：

- **資料品質和預先處理** – 確定用於訓練模型的資料具有高品質，且沒有錯誤、遺失值和不一致。資料預先處理步驟，例如處理遺失值、極端值偵測和特徵工程，在改善模型準確性方面扮演重要角色。
- **足夠的歷史資料** – 擁有足夠的歷史資料對於擷取模式、趨勢和季節性至關重要。不過，考慮歷史資料的相關性和及時性也很重要。如果市場、業務營運或外部因素發生重大變更，較舊的資料可能無法代表目前的情況。在這種情況下，將更高的權重提供給最近的資料。
- **功能選擇和工程** – 從現有資料識別相關功能和工程新功能可大幅改善模型的效能。與網域專家緊密合作，在選擇適當的功能時運用他們的知識和洞察。此外，請考慮執行功能重要性分析，以識別最具影響力的功能，並可能移除冗餘或不相關的功能。
- **整合模型** – 考慮使用結合多個模型預測的整合技術，而不是依賴單一模型。整合模型可以超越個別模型，並提供更強大且準確的預測。
- **模型評估和驗證** – 使用適當的指標定期評估和驗證模型的效能，例如均方誤差 (MSE)、均方絕對百分比誤差 (MAPE) 或任何其他特定網域的指標。使用交叉驗證或保留驗證來評估模型的概括化功能。
- **持續監控和重新訓練** – 貨運需求模式可能會因為各種因素而隨時間變化，例如經濟條件、市場動態或業務營運的變化。持續監控模型的效能，並使用最新的資料定期重新訓練模型，以提高其準確性和相關性。
- **可解釋的 AI** – 需求預測模型應該是可解釋和可解釋的，特別是在利益相關者需要了解預測背後原因的情況下。特徵重要性分析、部分相依性圖、Shapley 附加說明 (SHAP) 等技術可協助解釋模型的決策。
- **整合領域知識** – 與領域專家和業務利益相關者緊密合作，將他們的知識和洞察納入建模程序。他們的領域專業知識有助於識別潛在的偏差、解釋結果，並根據預測做出明智的決策。
- **案例分析和假設模擬** – 將執行案例分析和模擬的能力納入預測解決方案。這可讓利益相關者探索不同業務決策或外部因素對需求預測的影響，從而做出更明智的決策。
- **自動化且可擴展的管道** – 建置自動化且可擴展的管道，用於資料擷取、預先處理、模型訓練和部署。這會持續且有效率地執行預測程序，特別是在處理多個產品或區域時。

後續步驟

在實作此需求預測解決方案之前 AWS，建議您評估嘗試解決的問題。最好將企業擁有者和資料科學家集合在一起，以腦力激盪出 ML 模型是否可以解決問題。請務必了解您擁有哪些資料集，以及歷史資料的長度。企業擁有者也請務必與資料科學家合作，以提供網域知識、識別實用功能，並協助建立這些功能。模型的可靠性隨著您可以建立的相關功能數量而增加，這可提供更準確的預測。

若要建置此架構 AWS，請先設定 AWS 帳戶 並佈建必要的服務，例如用於資料儲存的 Amazon Simple Storage Service (Amazon S3)，以及用於機器學習模型訓練的 Amazon SageMaker AI。接下來，識別並收集將用作預測模型輸入功能的內部和外部資料來源。將此資料儲存在 Amazon S3 中，並使用 SageMaker AI 中的資料處理功能來預先處理和準備資料以進行模型訓練。在 SageMaker AI 中，使用自動模型調校和分散式訓練功能來訓練和最佳化預測模型。您也可以使用 AWS 服務 例如 AWS Step Functions 或 AWS Lambda 來設定管道，以使用最新的資料定期重新訓練預測模型。重新訓練後，在 SageMaker AI 中啟動批次轉換任務，以產生您存放在 Amazon S3 中的預測結果。使用 Amazon Quick 視覺化和監控從批次轉換任務產生的預測結果。

Resources

Amazon Quick 文件

- [什麼是 Amazon Quick ?](#)
- [使用資料來源](#)

Amazon SageMaker AI 文件

- [什麼是 Amazon SageMaker AI ?](#)
- [準備資料](#)
- [執行自動模型調校](#)
- [使用批次轉換](#)

AWS 程式碼範例

- [Amazon SageMaker AI 範例儲存庫](#) (GitHub)

文件歷史紀錄

下表描述了本指南的重大變更。如果您想要收到未來更新的通知，您可以訂閱[RSS摘要](#)。

變更	描述	日期
初次出版	—	2024 年 5 月 14 日

本文為英文版的機器翻譯版本，如內容有任何歧義或不一致之處，概以英文版為準。