Statistics and Application 统计学与应用, 2024, 13(3), 722-729
Published Online June 2024 in Hans. https://www.hanspub.org/journal/sa
https://doi.org/10.12677/sa.2024.133072

Hans汉斯

基于深度学习方法下的产品订单的数据分析与需求预测

巢　逸

浙江财经大学数据科学学院，浙江　杭州

收稿日期：2024年5月29日；录用日期：2024年6月18日；发布日期：2024年6月27日

摘　要

近年来企业外部环境变化不确定，企业供应链面临着更多的挑战。因此，需求预测作为企业供应链中的第一道防线，能够帮助企业更好地制定采购计划和生产计划，减少业务波动对企业的影响。针对销售区域、销售时段与订单需求量的关系分析，本文采取了Prophet模型。该模型能够捕捉到销售区域和销售时段对订单需求量的影响，并对现有的和未来的订单需求量进行准确的预测。针对订单需求量的整体影响，本文采用了GNN模型，考虑了销售区域编码、产品编码、产品大类编码、产品细类编码、销售渠道名称、产品价格等多种因素。该模型能够对订单需求量进行全面准确的预测，分析出影响因素占比大小关系，为企业提供决策支持。综上所述，本文的研究结果能够帮助企业更准确地预测未来的需求量，制定更好的采购计划和生产计划，从而提高企业的运营效率和竞争力。

关键词

需求预测，Prophet模型，GNN模型，经营策略

Data Analysis and Demand Forecasting for Product Orders Based on Deep Learning Methods

Yi Chao

School of Data Science, Zhejiang University of Finance and Economics, Hangzhou Zhejiang

Received: May 29th, 2024; accepted: Jun. 18th, 2024; published: Jun. 27th, 2024

Abstract

In recent years, the uncertainty of changes in the external environment has posed greater chal-

文章引用: 巢逸. 基于深度学习方法下的产品订单的数据分析与需求预测[J]. 统计学与应用, 2024, 13(3): 722-729.
DOI: 10.12677/sa.2024.133072

lenges to corporate supply chains. Therefore, demand forecasting, as the first line of defense in a company's supply chain, can help enterprises better formulate procurement and production plans, thereby reducing the impact of business fluctuations. This paper employs the Prophet model to analyze the relationship between sales regions, sales periods, and order demand. The model captures the influence of sales regions and periods on order demand and provides accurate forecasts for both current and future order demand. For the overall impact on order demand, this paper utilizes the GNN model, taking into account various factors such as sales region codes, product codes, major product categories, sub-product categories, sales channels, and product prices. This model enables a comprehensive and accurate prediction of order demand, analyzing the relative significance of influencing factors to support corporate decision-making. In conclusion, the research findings of this paper can help enterprises more accurately forecast future demand, develop better procurement and production plans, and thereby improve operational efficiency and competitiveness.

Keywords

Demand Forecasting, Prophet Model, GNN Model, Business Strategy

Copyright © 2024 by author(s) and Hans Publishers Inc.
This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).
http://creativecommons.org/licenses/by/4.0/
Open Access

1. 引言

近年来，随着大数据技术的不断发展和普及，数据挖掘成为了企业提高运营效率和竞争力的重要手段之一。在制造业中，面向经销商的出货数据是一种重要的信息来源，对于制造企业来说，了解自己产品在市场上的销售情况和受欢迎程度是十分必要的。

在这个背景下，该企业提供的面向经销商的出货数据就显得尤为重要。这份数据反映了该企业产品在不同销售区域的价格和需求等信息，是制造企业了解产品销售情况的重要依据。通过对这份数据的挖掘和分析，制造企业可以了解到产品在不同销售区域的受欢迎程度，产品价格对需求量的影响，以及不同销售方式、不同品类、不同时间段和节假日等因素对产品销售的影响，为企业制定销售策略和优化产品结构提供有力支持。

数据挖掘的重要性不仅在于对企业决策的帮助，也在于提高企业的效率和竞争力。通过数据挖掘，制造企业可以更好地了解市场需求和客户需求，从而更好地优化产品结构，提高产品质量，提升品牌价值，进而提高企业的市场占有率和盈利能力。此外，数据挖掘还可以帮助企业优化供应链管理、降低成本和提高生产效率，从而实现企业可持续发展。刘晶(2016) [1]针对线上农产品销售存在的信息不对称问题，提出一种结合深度学习算法，实现线上农产品的销量分类预测。李同欢(2017) [2]提出一种基于深度学习的多交互混合推荐模型，通过深度学习模型融合更多的辅助信息作为输入，能够缓解数据的稀疏性问题。Olofintuyi Sunda 等(2023) [3]提出了一种使用卷积神经网络和循环神经网络(CNN-RNN)与长短期记忆(LSTM)相结合的深度学习方法来进行可可产量预测。

因此，对于制造企业来说，数据挖掘已经成为了不可或缺的一部分。通过对经销商的出货数据的挖掘和分析，制造企业可以更好地了解市场需求和客户需求，提高企业的效率和竞争力，为企业的发展提供强有力的支持。

2. Prophet 模型下销售区域、销售时段与订单需求量的关系分析

Prophet 是 Facebook 开源的一种时间序列预测模型，它可以自动拟合季节性变化、节假日效应以及异常值等特征，从而准确地预测未来的需求趋势和波动。所以在探究销售区域、销售时段与订单需求量的关系分析时采用 Prophet 模型。

2.1. 数据处理

在设计 Prophet 模型前，还是沿用 MLP 感知器模型下数据预处理的方法对数据进行清洗，包括数据清洗：缺失值处理，异常值处理等；将数据按照时间序列排序，将订单日期转换为日期格式等从而去除异常值对模型产生的影响。

2.2. 模型的选择与训练

Prophet 模型使用一个可分解的时间序列模型，主要由趋势项(trend)，季节项(seasonality)和假期因素(holidays)组成：

$$y(t)=g(t)+s(t)+h(t)+\varepsilon_t$$

其中 $y(t)$ 是趋势函数，代表非周期变化的值，$g(t)$ 表示周期性变化(如每周和每年的季节性)，$h(t)$ 表示在可能不规律的时间表上发生的假期的影响。ε_t 误差项代表模型不能适应的任何特殊变化，并假设其符合正态分布。Prophet 实现了两个趋势模型，分别是基于逻辑回归的饱和增长模型和分段线性模型。首先是基于逻辑回归的趋势项：

$$g(t)=\frac{C}{1+\exp\left(-k(t-m)\right)}$$

C 为承载力，k 为增长速率，m 为偏移参数。

其次是模型为分段逻辑回归模型：

$$g(t)=\frac{C(t)}{1+\exp\left(-\left(k+\alpha(t)^t\delta\right)\cdot\left(t-\left(m+\alpha(t)^T\right)\gamma\right)\right)}$$

2.3. 基于 Prophet 模型下销售区域、销售时段的订单供求关系分析

2.3.1. Prophet 模型的设计和参数设置

首先，Prophet 模型可以拟合数据中的季节性趋势、节假日效应等因素，并预测未来的需求量，因此我们可以使用 Prophet 模型来预测不同销售区域在不同时间段的订单需求量。在设计 Prophet 模型时，需要考虑不同的参数来控制模型的性能和拟合效果，包括：季节性参数、趋势性参数和非线性参数等。具体流程如下：首先进行数据预处理，去除部分缺失值；其次，根据确认先验知识是否存在确定节假日及其规模并判断数据周期数据；然后，根据趋势向周期以及节假日附加项进行预测值修正，最后模型接收参差数据，选择合适参数，输出迭代结果。

2.3.2. Prophet 模型下同一区域不同时间段产品需求量变化

现探究在同一区域(101~105)下不同时间段产品需求量变化情况，在使用 Prophet 模型时会有一定的预测效果，其中各地区产品需求量拟合模型的 RMSE(均方根误差)结果如表 1 所示。

从结果中可以看出，101~105 区域整体销售量都位于拟合范围内即蓝色阴影部分中，整体数据都靠近历史数据点，只有极少部分数据位于拟合范围外，不排除其他因素的干扰导致某天或者某月销售量出

现异常情况。对于 5 个区域未来 90 天预测情况来看，101、102、103 和 105 区域产品销售量会出现先上升后下降趋势，其中 103 地区产品销售量上升幅度最大，但整体都会处于估计的区间范围内；104 地区产品销售量再过去的时间段中出现持续上升情况，未来区域变化犹未可知。

Table 1. RMSE values for different regions under the prophet model
表 1. Prophet 模型下不同区域 RMSE 值

区域	101	102	103	104	105
RMSE	60.83	58.94	65.23	78.46	67.31

2.4. Prophet 模型下同一时间段不同区域产品需求量变化

对于同一时间段下不同区域产品需求量变化图如下：在所给的时间段(2015 年到 2018 年年底)中，各区域(101~105)总体波动即各地区的产品销量整体数量相同，处于一定范围内的波动，说明该产品在各地区总体销售量差不多，其中 105 地区产品销售量在 2015 年刚开始的销售量最好，但是在每年中旬时产品销售量达到最低谷，此后一直处于平稳波动状态。

2.5. Prophet 模型下探索节假日对产品需求量影响

为了探求节假日对产品销售量的影响，本文选取元旦、春节、劳动节和国庆节这些代表性节日来探究对销量的影响。然后，我们创建了 Prophet 模型对象，并将这些节假日效应添加到模型中。

第一部分趋势图展示了销售量的整体趋势和预测值。在该图中，蓝色线表示历史销售数据的趋势，黑色线表示历史和未来的销售数据预测趋势，蓝色区域表示历史数据的不确定性，黑色区域表示历史和未来预测数据的不确定性。可以看到，再过去的时间段里，销售量一直处于平稳状态，未来的销售量预测趋势变化的不确定性也会很小。

第二部分节假日效应图显示了定义的节假日(元旦、春节、劳动节和国庆节)对销售量的影响。在该图中，可以看到每个节假日都有一个相应的效应图，其中重叠线表示节日发生的时间，蓝色区域表示节日效应的时间窗口。在这个例子中，每个节假日对应的效应在节日前后都有所增加，也即是两根实线对比高度的变化，它们的变化差值也即是在节假日效应下，产量的增减变化情况。即在节假日的效应下，产品的销售量会出现增长。

第三部分季节性效应图展示了时间序列数据中的周期性变化，包括每年和每周的季节性变化。在这张图中，可以看到每周的销售量呈现出明显的周期性变化，每年的销售量也呈现出季节性的波动，这种波动往往与季节性的节日相关，例如每年的春节、劳动节、国庆节等。在周的季节性变化中，双休日是销售量最高的时候，而周二则是产品销售量最低的时候。

2.6. Prophet 模型下企业未来相关调整措施

根据上述结论，企业可以考虑在产品营销策略上做出相应的调整。首先，由于双休日是销售量最高的时候，可以适当在周末加强促销活动，以刺激顾客消费欲望，提高产品销售量。其次，由于周二是产品销售量最低的时候，可以考虑在周二推出特别促销，以吸引顾客注意，提高产品销售量。此外，在节假日等特殊时期，也需要根据节假日效应进行相应调整，以更好地适应市场需求，提高产品竞争力。例如，在 101、102 和 105 地区，可以考虑加大市场营销力度，增加品牌曝光度，以提高产品知名度和消费者对产品的信任感，从而提升销售量。而在 103 和 104 地区，则可以重点推广销售量较高的产品，提高产品的市场占有率，进一步扩大销售规模。此外，企业还可以根据周的季节性变化，在销售淡季采取促销活动等措施，以提高销售量。结合以上分析，可以制定更为精准的产品营销策略，增强企业市场竞争

力，进一步提高企业的盈利能力。

3. 基于 GNN 模型下销售区域与指标影响分析

3.1. GNN 模型的设计和参数设置

3.1.1. GNN 模型的介绍

GNN 是一种基于图结构数据的神经网络，常用于处理复杂的非欧几里得数据，如社交网络、蛋白质分子结构等。在本例中，可以将每个销售区域编码作为一个节点，将同一区域内的订单数据看做节点之间的边，构建成一个图结构数据，然后使用 GNN 对这个图进行建模，从而探究不同指标对销量的影响。

3.1.2. 指标影响分析下 GNN 模型的参数设置

具体来说，可以通过将每个订单的其他特征(如产品编码、销售渠道名称等)作为节点的属性，并将订单需求量作为边的权重，构建一个带权有向图。然后使用 GNN 对这个图进行建模，以预测每个销售区域的销量。

GNN 的目的就是为每个节点学习到一个状态嵌入向量 $h_v \in R^s$，这个向量包含每个节点的邻居节点的信息，h_v 表示节点的状态向量，这个向量可以用于产生输出 R^s，假设存在局部转移函数 $f(\cdot)$，这个函数在所有节点中共享，并根据邻居节点的输入来更新节点状态，$g(\cdot)$ 为局部输出函数，这个函数用于描述输出的产生方式，则：

$$h_v = f\left(x_v, x_{co[v]}, h_{ne[v]}, x_{ne[v]}\right)$$

$$o_v = g\left(h_v, x_v\right)$$

其中 x_v 表示节点 v 的特征向量，$x_{co[v]}$ 表示与节点 v 关联的边的特征向量，$h_{ne[v]}$ 表示节点 v 的邻居节点的状态向量，$x_{ne[v]}$ 表示节点 v 的邻居节点特征向量。

假设将所有的状态向量，所有的输出向量，所有的特征向量和所有的节点特征而得到的向量叠加起来，分别用 H、O、X、X_N 表示，则：

$$H = F\left(H, X\right)$$

$$O = G\left(H, X_N\right)$$

F、G 分别称为全局转移函数和全局输出函数，是图中针对所有节点的 f、g 的堆叠版本，根据 Banach 的不动点定理，GNN 使用如下的传统迭代方法来计算状态参量：

$$H^{t+1} = F\left(H^t, X\right)$$

其中 H^{t+1} 是 H 的第 t 个迭代周期的张量。

3.2. GNN 模型下销售量影响因子大小关系

3.2.1. 模型评估与调优

这个模型的 mse 值包括训练集 mse 和验证集 mse 两个部分，随着 epoch 的增加，两个部分的 mse 值都呈现下降趋势，表明模型在不断学习和提高。当 epoch 值增加到 12 时，验证集 mse 值趋近于平稳且较小，达到 0.004，这意味着模型已经接近最优状态，并且具有一定的泛化能力。同时，模型的 loss 值图和 mse 图呈现相似的下降趋势，这表明模型的学习过程相对稳定，损失函数和评价指标的变化趋势一致。在模型评估过程中，除了 mse 值，还可以考虑其他评价指标，如准确率、召回率等以综合评估模型的性能，并进行调优。

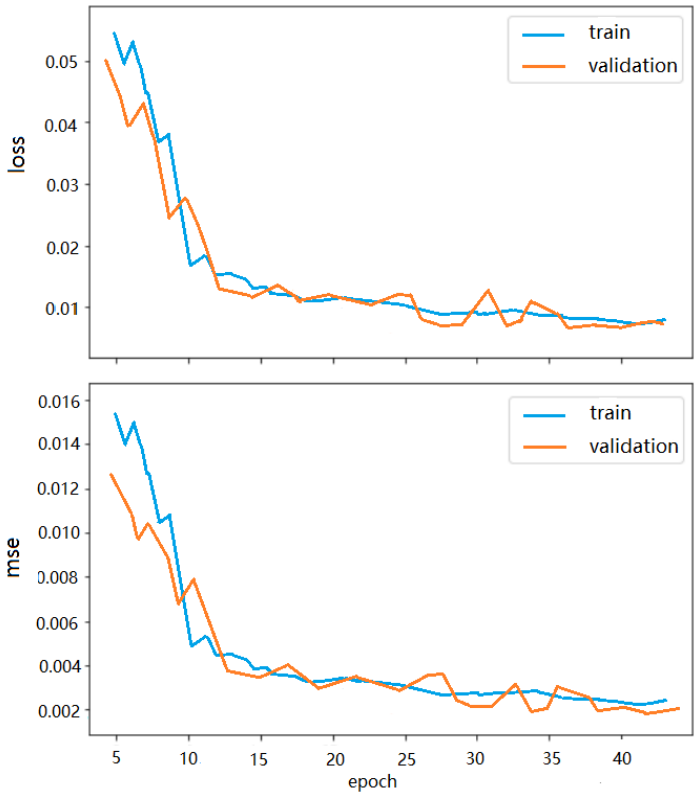

Figure 1. Training status of the GNN model
图 1. GNN 模型训练情况

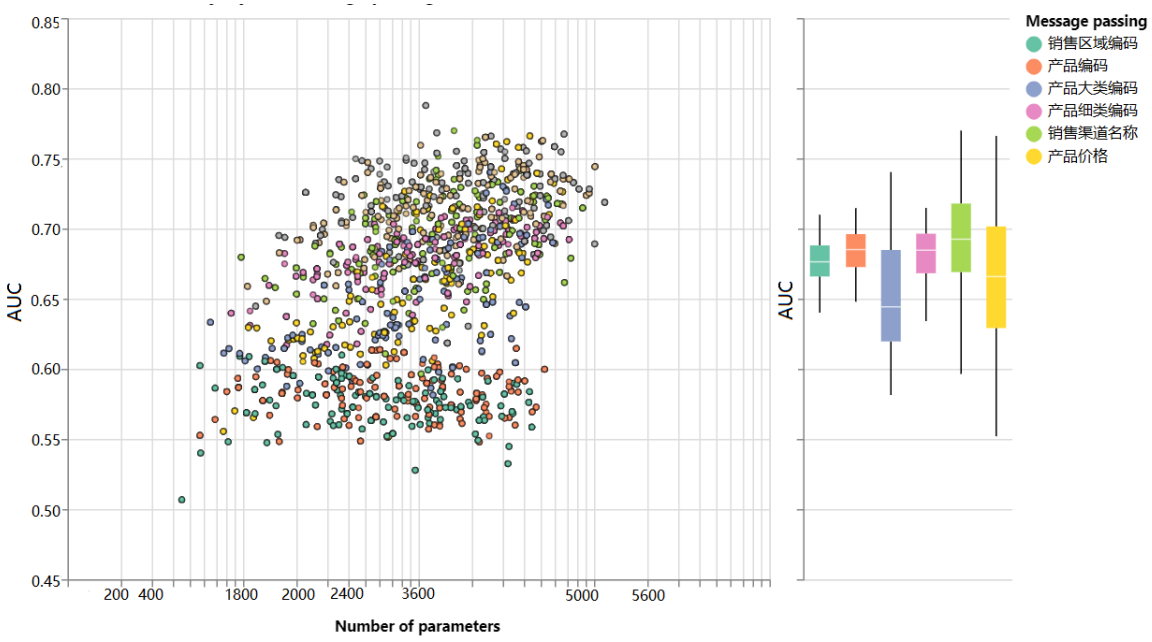

Figure 2. Distribution of AUC scores for different variables under the GNN model
图 2. GNN 模型下不同变量的 AUC 分布情况

其中 AUC 可视化图是一种常用于分类模型评估的指标，表示分类器在所有阈值下的性能综合，在

GNN 模型中，AUC 可视化图主要用于评估模型的分类能力，即模型对正负样本的区分能力。通过观察 AUC 可视化图的形状和数值，可以对模型的性能进行评估和比较。

图 2 显示所选取的指标取值范围都在 0.5 到 1 之间，数值越接近 1 则表示分类器的性能越好，所选取的指标都可以反映出销售量的变化关系。

3.2.2. 销售量影响因子大小关系

为了探究出对于销售量各影响因素占比大小关系，在 GNN 模型框架下，对各影响因素进行可视化结果分析。

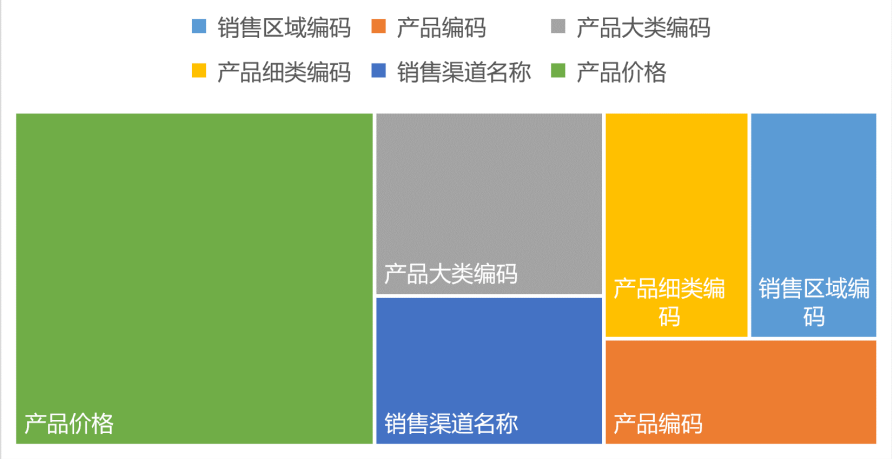

Figure 3. Variable importance in the GNN model
图 3. GNN 模型下的变量重要性

从图 3 可视化结果可以看出，在预测销售量的过程中，产品价格是最重要的影响因素(−0.12)。这可能是因为价格是顾客在购买时最关注的因素之一，因此价格高低对于销售量的影响非常明显。其次，产品大类编码和销售渠道名称对于销售量的影响也非常重要。产品大类编码可以反映出不同类别产品之间的销售量差异，而销售渠道名称则反映出不同销售渠道之间的差异，这对于制定销售策略具有重要的参考意义。

然而，产品编码对于销售量的影响较小，这可能是因为产品编码本身并不是消费者在购买时所关注的重要因素，而更多的是为了管理和跟踪产品的生命周期。因此，在销售策略的制定中，不应将过多的精力和资源投入到产品编码方面。

此外，通过 GNN 模型，我们还可以看出不同的因素之间的相互影响。例如，各影响因素都会对销售量产生一定的影响，只是大小关系不同；产品价格和销售渠道名称之间存在相互影响，这表明选择不同的销售渠道也会影响产品的价格策略。这为制定更加全面的销售策略提供了有力的参考。

3.3. GNN 模型下企业未来相关调整措施

基于 GNN 模型的分析结果，企业可以采取以下措施来调整未来的销售策略：

价格优化：由于产品价格对销售量影响最大，企业可以通过合理定价来提高销售量。调整销售渠道：销售渠道名称也是影响销售量的一个重要因素，因此企业可以考虑调整销售渠道，加强线上渠道建设，提高电商平台的曝光度和销售效率；产品分类管理：产品大类编码也对销售量产生了影响，因此企业可以通过对不同大类产品的管理和分类，针对不同产品采用不同的销售策略，以提高销售量。综上，GNN

模型对于评判指标影响因素大小关系非常有用。如果企业能够预测和利用 GNN 变化，就能更好地满足消费者的需求，增加销售额，提高企业的竞争力。

4. 结论与建议

在销售策略方面，我们提出了以下建议：

1. 周末是顾客消费的高峰期，因此加强促销活动可以吸引更多的顾客到店消费，提高产品销售量。可以采取一些具有吸引力的促销手段，例如打折、赠品、抽奖等，以激发顾客购买欲望，提高产品销售量。

2. 周二通常是顾客消费的较低谷期，推出特别促销可以吸引顾客的注意，提高产品销售量。可以选择一些具有特色的产品进行推广，并采取一些优惠措施，如限时折扣、买一送一等，以吸引顾客到店消费。

3. 节假日对消费者的购买行为有很大的影响，因此根据节假日效应进行相应调整可以更好地适应市场需求，提高产品竞争力。可以根据不同的节假日特点，设计相应的促销活动或推出符合节日主题的产品，以吸引顾客的购买欲望，提高产品销售量。

4. 在 101、102 和 105 地区，可以考虑加大市场营销力度，增加品牌曝光度，以提高产品知名度和消费者对产品的信任感，从而提升销售量。可以选择一些有影响力的媒体进行广告投放，或与当地的商会、行业协会合作开展宣传活动，提高产品的知名度和美誉度，从而吸引更多的顾客购买。

5. 在 103 和 104 地区，则可以重点推广销售量较高的产品，提高产品的市场占有率，进一步扩大销售规模。可以根据不同产品的销售情况，选择一些销售量较高的产品进行重点推广，并采取一些优惠措施，如打折、赠品等，以提高产品的市场占有率和销售量。

6. 销售淡季是产品销售的低谷期，采取促销活动等措施可以提高销售量。可以选择一些有吸引力的促销手段，例如限时打折、买一送一等，以激发顾客购买欲望，提高产品销售量。

7. 产品价格对销售量的影响最大。在定价时，应根据产品的不同价格区间采取不同的定价策略以提高销售量。同时，还可以结合其他因素，如促销活动、季节变化等，制定更加科学的销售策略，提高产品竞争力。

总的来说，了解不同销售区域的需求差异，了解季节性和节假日的影响，关注不同产品系列的价格敏感性，是企业制定有效销售策略并实现可持续销售增长的关键因素。

参考文献

[1] 刘晶, 和述群, 朱清香, 等. 基于深度学习的线上农产品销量预测模型研究[J]. 计算机应用研究, 2017, 34(8): 2291-2293, 2344.

[2] 李同欢, 唐雁, 刘冰. 基于深度学习的多交互混合推荐模型[J]. 计算机工程与应用, 2017, 55(1): 135-141.

[3] Olofintuyi, S.S., *et al*. (2023) An Ensemble Deep Learning Approach for Predicting Cocoa Yield. *Heliyon*, **9**, e15245. https://doi.org/10.1016/j.heliyon.2023.e15245