

基于XGBoost模型产品订单量的预测

苏邳宏, 孙宇婷, 杨永铖, 廖好婕, 黄航英, 金秀玲*

闽江学院, 数学与数据科学学院(软件学院), 福建 福州

收稿日期: 2023年9月12日; 录用日期: 2023年11月2日; 发布日期: 2023年11月9日

摘要

把握供应链需求, 提高资源配置和利用效率对企业树立竞争优势具有重大的实际意义。通过特征分析提取了特征变量后, 建立XGBoost模型, 利用GridSearchCV技术对模型进行调参优化, 最终得到模型的拟合优度为0.956, 均方根误差RMSE为36.522。同时我们还建立了随机森林模型进行对比, 结果表明XGBoost模型效果更理想, 因此最终根据所构建的XGBoost模型实现对企业未来三月产品订单量的预测, 为企业合理安排生产计划提供了一定的理论依据。

关键词

产品订单量, XGBoost模型, 随机森林, 预测

Prediction of Product Order Volume Based on XGBoost Model

Zhihong Su, Yuting Sun, Yongcheng Yang, Yujie Liao, Hangying Huang, Xiuling Jin*

College of Mathematics and Data Science (Software College), Minjiang University, Fuzhou Fujian

Received: Sep. 12th, 2023; accepted: Nov. 2nd, 2023; published: Nov. 9th, 2023

Abstract

Grasping supply chain demand, improving resource allocation and utilization efficiency is of great practical significance for enterprises to establish competitive advantage. After extracting characteristic variables through feature analysis, XGBoost model is established, and the model is adjusted and optimized using GridSearchCV technology. Finally, the goodness of fit of the model is 0.956, and the Root-mean-square deviation RMSE is 36.522. At the same time, we also established

*通讯作者。

a Random forest model for comparison. The results indicate that the XGBoost model has a more ideal effect. Therefore, based on the constructed XGBoost model, the prediction of the company's product order volume in the next three months is ultimately achieved, providing a certain theoretical basis for the company to reasonably arrange production plans.

Keywords

Product Order Volume, XGBoost Model, Random Forest, Prediction

Copyright © 2023 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

随着数字化经济的不断发展, 电子商务在全球范围内呈现迅猛的发展趋势, 成为推动经济增长和转型升级的重要力量。而近年来随着企业内外各种因素不断增多, 对产品需求的预测难度增加。作为现代企业管理中不可或缺的一环, 产品需求量预测对于企业提升竞争力和市场份额具有至关重要的作用。企业若能准确地预测未来的需求量, 就能减少库存积压、缺货以及运营成本的浪费, 提高生产效率和客户满意度, 进一步提升企业的市场竞争力。

国内外学者对于订单需求预测类研究主要以时间序列模型为基础, 其中国外学者 W.J.Kim [1]利用传统的 ARIMA 模型对于施工订单量进行预测分析; M.Khashei、Voronin 等[2] [3]人比较了线性与非线性模型对订单预测的优势, 将 ARIMA 模型与 ANN 模型混合, 提升了预测精度。在国内的相关研究中, 王斌[4]等人通过构建 ARIMA 模型来实现对新疆灰枣订单量预测, 及时把控了市场的供求关系; 谭祖健[5]等人运用 ARIMA 模型的滑动平均功能以及自回归功能来对发动机订单需求进行分析预测; 施佳[6]基于 ARIMA-BP 组合模型, 结合了线性与非线性的特征来实现餐饮企业订单预测; 欧红森[7]等人提出了 GA-BP 神经网络模型来实现对船舶备件的需求预测。

综上所述, 对于现有的订单需求预测方法虽然具有一定的合理性, 但仍存在一些问题。对于传统方法预测订单需求时, 较为依赖历史数据, 数据缺失或者记录不准确会影响预测结果的准确性[8], 并且对于非线性变化的反应相对迟钝; 其他预测方法虽然进行了一定的改进, 但是仍存在信息提取不完备的现象。考虑到集成学习能够将多个弱分类器集合为一种强分类器, 不仅容易陷入局部最优解, 还可以处理大量、多样化、非线性的数据, 速度快、准确率高, 能够有效地解决传统模型带来的问题。因此, 本文通过构建 XGBoost 模型, 以期实现对用户订单需求的准确预测, 进而帮助企业适应变化的市场环境, 调整策略[9]。

2. 研究方法

2.1. XGBoost 算法原理[10] [11]

XGBoost (Extreme Gradient Boosting)算法是一种梯度提升决策树(Gradient Boosting Decision Tree, GBDT)的变种, 属于一种集成学习方法。

1) XGBoost 模型结构: XGBoost 采用了 CART 树作为基分类器。在 GBDT 的基础上, XGBoost 在每一轮迭代中引入了正则化项, 以控制模型的复杂度, 并且支持 L1 和 L2 正则化。

2) **XGBoost 损失函数**: 对于回归问题, XGBoost 采用平方损失函数, 对于二分类问题, 采用对数损失函数。在每一轮迭代中, XGBoost 会计算样本当前条件下梯度和 Hessian 矩阵的值, 并通过迭代优化损失函数来拟合目标值。

3) **XGBoost 提升策略**: 与 GBDT 不同, XGBoost 采用了类似 Adaboost 的加权平均策略。在训练过程中, 每个样本都会被赋予一个权重, 在下一轮迭代时, 会更加关注那些之前分类错误的样本。

XGBoost 算法在产品订单的需求预测任务中具有很大的优势。首先, 它可以以高效地处理大规模的数据集, 对于产品订单数量较多的情况, 帮助我们提高模型训练和预测的效率。其次, 它在预测准确度上表现出色, 可以处理多个变量之间的复杂关系, 对于产品订单需求预测这样的复杂任务, 能够提供更加准确和鲁棒的预测结果。此外, 还具有较强的可解释性, 可以输出各变量的重要性和贡献值, 帮助我们更好地理解模型预测结果的形成过程。最后, XGBoost 算法还可以进行正则化和并行计算, 避免过拟合和提高计算效率, 更好地应对产品订单需求预测的实际场景。综上, XGBoost 算法在进行产品订单的需求预测时具有一定的优势, 可以帮助企业更好地把握市场需求的变化, 优化订单规划和生产调度的决策。

2.2. 模型评价标准

采用自定义评估损失函数来评估模型的指标, 根据均方根误差(RMSE)、平均绝对误差(MAE)、对称平均绝对百分比误差(SMAPE)以及拟合优度(R^2)来作为模型预测精度的评估标准[12]。因为 RMSE 可以评价预测值和真实值之间平均偏差的指标, 当产品订单预测值与真实值相差较大时, RMSE 可以帮助我们发现问题并进行调整, 从而提高预测的准确性; MAE 可以反映出预测值和真实值之间的绝对误差平均数, 其注重模型对于异常值的拟合情况; SMAPE 是计算预测值和真实值之间的百分比误差的平均值, 能够更好地反映产品订单预测中需求量相差较大的情况, 例如生产季节性需求的预测, 涉及到季节性变化的货物需求量变化; R^2 则能更好的反映线性回归预测值与真实值的拟合程度。具体公式如下:

均方根误差:

$$RMSE = \sqrt{\frac{1}{N} \sum_{t=1}^N (\text{observed}_t - \text{predicted}_t)^2}$$

平均绝对误差:

$$MAE(X, h) = \frac{1}{m} \sum_{i=1}^m |h(x^{(i)}) - y^{(i)}|$$

对称平均绝对百分比误差:

$$SMAPE = \frac{100\%}{n} \sum_{i=1}^n \frac{|\hat{y}_i - y_i|}{(|\hat{y}_i| + |y_i|) \div 2}$$

拟合优度:

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y}_i)^2}$$

本文的基本研究框架如图 1 所示。

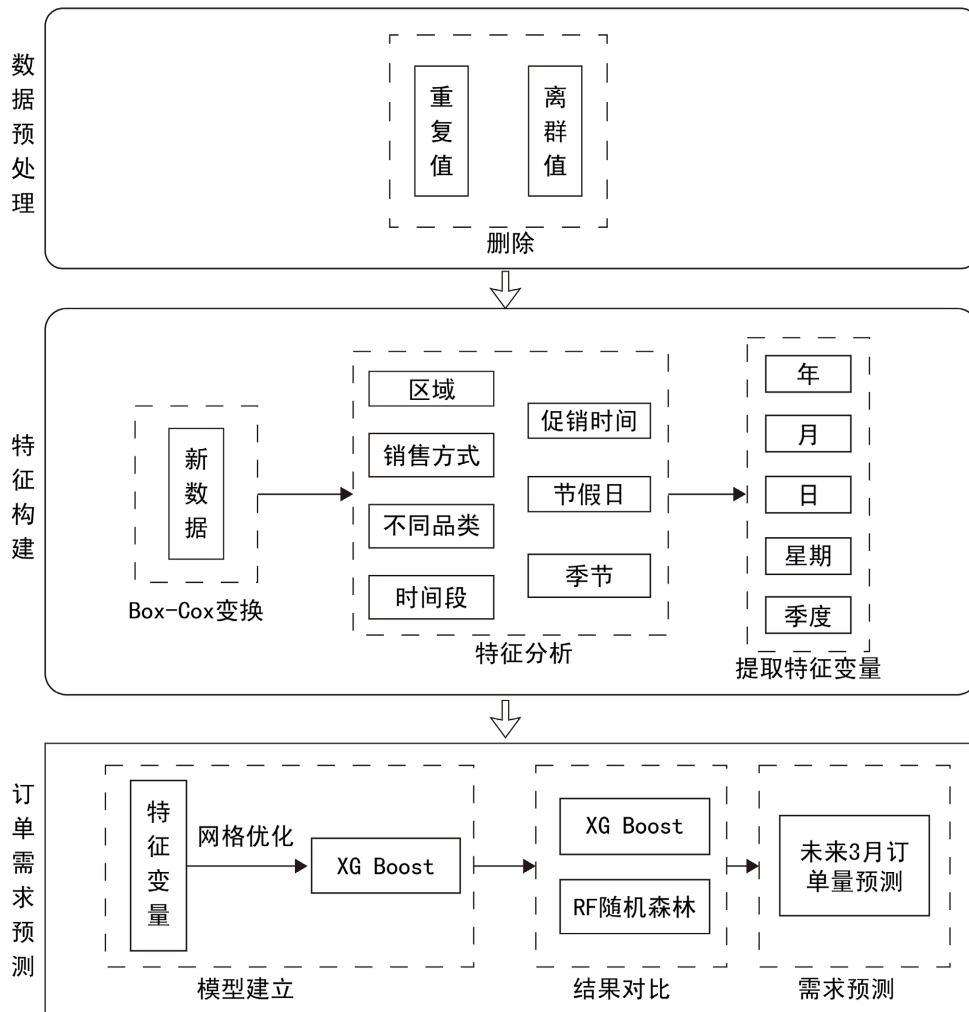


Figure 1. Basic research framework
图 1. 基本研究框架

3. XGBoost 预测产品订单量

3.1. 数据来源

文章的数据来源于第十一届泰迪杯 B 题数据，是关于我国某大型企业的面向经销商的出货数据，数据真实可靠。其时间跨度为 2015/09/01~2018/12/20，共有 8 个变量，597,694 行，分别为 order_date (订单日期)、sales_region_code (销售区域编码)、item_code (产品编码)、first_cate_code (产品大类编码)、second_cate_code (产品细类编码)、sales_chan_name (销售渠道名称)、item_price (产品价格)和 ord_qty (订单需求量)。

3.2. 数据预处理

原始数据中存在缺失值的情况，首先删除了多余的重复数据，再根据箱线图的原理进行异常值处理，将产品价格按照产品编码进行分组，对产品价格中小于下界或大于上界的离群点进行索引并剔除。最终删除的离群点共有 47,143 个，处理后的数据集中共有 550,551 行。原始数据前 8 行预览如表 1 所示。

Table 1. Preview of the first 8 rows of original data
表 1. 原始数据前 8 行预览

order_date	sales_region_code	item_code	first_cate_code	second_cate_code	sales_channel_name	item_price	ord_qty
2015-9-1	104	20,028	301	405	offline	1012	12
2015-9-1	104	22,069	307	403	offline	1114	19
2015-9-2	104	20,028	301	405	offline	1005	11
2015-9-2	102	20,323	305	412	offline	99	502
2015-9-2	104	20,448	308	404	online	962	3
2015-9-2	102	20,457	305	412	offline	164	308
2015-9-2	104	20,606	308	404	offline	1614	14
2015-9-2	104	20,606	308	404	offline	1615	17

3.3. 描述性分析

1) 价格对需求量的影响

通过对价格和需求量的组合分析，可以发现产品的不同价格对需求量的影响，基本呈正态分布价格上升需求量也会上升但会存在一个阈值点从而价格继续上升而需求量下降。

对价格与需求量进行相关性分析，其皮尔逊相关系数为-0.128，价格与需求量呈显著负相关，即价格越高，需求量越低，与上述描述性统计有所出入，说明还存在其他影响需求量的因素，需要进一步进行研究。

2) 产品所在区域对需求量的影响

通过差异性分析对不同区域之前需求量的差异情况进行研究，在实际分析过程中发现需求量不呈现正态性分布，故只能进行做多独立样本 Kruskal-Walis 检验，其检验结果 P 值小于 0.05，故说明不同区域在需求量上存在显著差异。进一步进行事后比较，发现 104 与其余四个地区、105 与 101、102 在需求量上存在显著差异。进一步研究各区域需求量间的特性，103 需求量均值最大，105 需求量中位数最小，104 需求量总量最小，但其需求量均值并不是最低，即其需求量总量最少主要是因为产品种类较少。

3) 不同销售方式(线上和线下)的产品需求量的特性

同种产品线上模式各类电商平台的需求量均值高于线下模式实体经销商。但各类型产品线上总量低于线下模式，考虑到是很多种类产品还没有开通线上售货渠道的原因。

4) 不同时间段产品需求量的特性

本文聚合了 12 个月的月初、月中、月末的订单需求量，通过每个时间段的平均需求量，总需求量和标准差进行分析，用来比较每个时间段内的订单需求量的分布情况，月中的平均高求量最大，而在总需求量方面，月末最高，是因为月末的产品种类需求更多。通过标准差，我们可以看出，月头月中月末基本持平，也就是说，其需求量序列基本都是平稳的。

5) 不同外因对产品需求量的影响

本文在研究了以上四类数据所提供的自变量外，进一步通过多独立样本检验研究了节假日、促销及季节因素对产品需求量的影响。检验结果见表 2，其中 1_1 为是节假日、2_1 为不是节假日，1_2 为是促销日、2_2 不是为促销日，1、2、3、4 分别代表春夏秋冬。

Table 2. Analysis results of multiple independent sample tests
表 2. 多独立样本检验分析结果表

变量名	变量值	中位数	标准差	P	Cohen's d 值
需求量	1_1	29	199.945	0	0.017
	2_1	26	189.093		
	1_2	29	2000.083	0	0.029
	2_2	26	171.531		
	1	27	226.990	0	0
	2	33	209.386		
	3	32	181.522		
	4	24	158.769		

从表 2 中可见，对于节假日因素，通过双独立样本检验可以得出，检验结果 P 值为 0，因此统计结果显著，故是否节假日在需求量上存在显著差异；其差异幅度 Cohen's d 值为 0.017，差异幅度非常小。并且节假日需求量中位数低于非节假日。

接着我们通过计算节假日订单需求量(monthly_holiday_order)与当月订单需求量(monthly_order)的比例，将得到的结果称为节假日订单需求量贡献度，经过分析得出，在这段时间内，节假日订单对当月订单的贡献度总体上呈现出一个规律性的变化，每年的元旦节与国庆节期间，对月订单需求量的贡献度较高。

对于促销因素，近些年促销日活动时间逐步拉长，为了更准确地探讨促销对产品需求的影响，本文研究了促销日的前七天和后七的天一段时间，分别为每年的 06.11~06.25 (618 促销)、11.04~11.18 (双 11 促销)、12.05~12.19 (双十二促销)。

计算促销数据和非促销数据的平均需求量，可以发现在所有年份中，促销期间的平均需求量大于非促销期间的平均需求量，这表明促销对需求产生了正面影响。通过促销日订单需求量增长率，得到结果：6.18 订单需求量增长率：263.92%、双 11 订单需求量增长率：351.41%、双 12 订单需求量增长率：358.17%。进一步说明促销可以使产品订单需求量呈现较大增长。

对于季节因素，将数据按照销售季节进行整合，计算每个季节的平均订单需求量，引用季节性指数来衡量时间序列中季节性变化对总体趋势的影响程度，通过计算不同季节的季节性指数来研究季节性对产品需求量的影响。Q1、Q2、Q3 和 Q4 的季节指数分别为 1.08、0.96、0.9、1.05，根据综合季节性指数可以得出：Q1 和 Q4 季节性指数较高，说明这两个季度产品的需求量相对较高；Q2 和 Q3 季节性指数较低，说明这两个季度产品的需求量相对较低。因此，季节因素对产品需求量有一定的影响。

4. 建立 XGBoost 模型预测产品订单量

4.1. 数据选取及模型训练

由于数据具有时序性，故选取全部数据的前 80%作为训练集，后 20%作为测试集。将销售区域编码，产品编码，产品大类编码，产品细类编码，销售渠道名称，年，月，日，星期，季度编码，季节编码，是否为月初，是否为月末，是否为季度初，是否为季度末，是否为年初，是否为年末，是否为节假日，价格区间编码和销售额共 20 个特征指标作为模型输入，将订单需求量作为模型输出，建立 XGBoost 模型进行训练，能够较为理想地预测产品的订单需求量。

模型训练参数设置如下：学习器个数为 1000，树最大深度为 5，学习率为 0.1，学习任务类型为 squarederror，随机种子为 42。将训练出来的模型对测试集上的数据进行预测，将预测结果与真实值进行拟合，并评估损失，评估结果得到 $R^2 = 0.955$ ， $RMSE = 36.713$ ， $MAE = 6.294$ ， $SMAPE = 23.317$ 。

4.2. 模型优化

我们采用近年来被广泛应用在机器学习模型优化的 GridSearchCV 网格搜索结合手动微调进行参数调校，GridSearchCV 是一种超参数调优技术，其原理是通过交叉验证从一个包含各种超参数选项的网格中找到最优的超参数组合，目的是能够保证最大程度上优化模型的性能又能防止过拟合。需要注意的是，由于 GridSearchCV 需要遍历整个超参数网格，所以可能会消耗大量时间和计算资源。因此，在实践中常常需要权衡精度和效率之间的平衡，并结合一些启发式方法来加速调参过程。

对训练集上的数据进行调参，经过网格搜索结合手动调校，得到最优参数组合如下：学习器个数为 1000，树最大深度为 8，学习率为 0.2，学习任务类型为 squarederror，随机种子为 42。优化后模型的 $R^2 = 0.956$ ， $RMSE = 36.552$ ， $MAE = 5.347$ ， $SMAPE = 16.718$ 。与原模型相比，MAE 减小了 0.947，SMAPE 减小了 6.599，模型性能显著提高。

4.3. 模型对比

将 XGBoost 模型与随机森林模型进行比较。建立随机森林模型(RF)进行训练，经过网格和手动调参后，选取到最优参数设置如下：学习器个数为 50，树最大深度为 35，最大特征数量为 sqrt，最小样本数量为 10，随机种子为 42。XGBoost 和 RF 在测试集上的拟合情况如图 2 所示，图中*代表 XGBoost 预测值数据点，△代表随机森林预测值数据点，另外在拟合时，我们只选取了预测数据与测试集的前 100 个数据进行拟合，因此图中横坐标表示 0~100 的等差数列，纵坐标表示订单需求量。

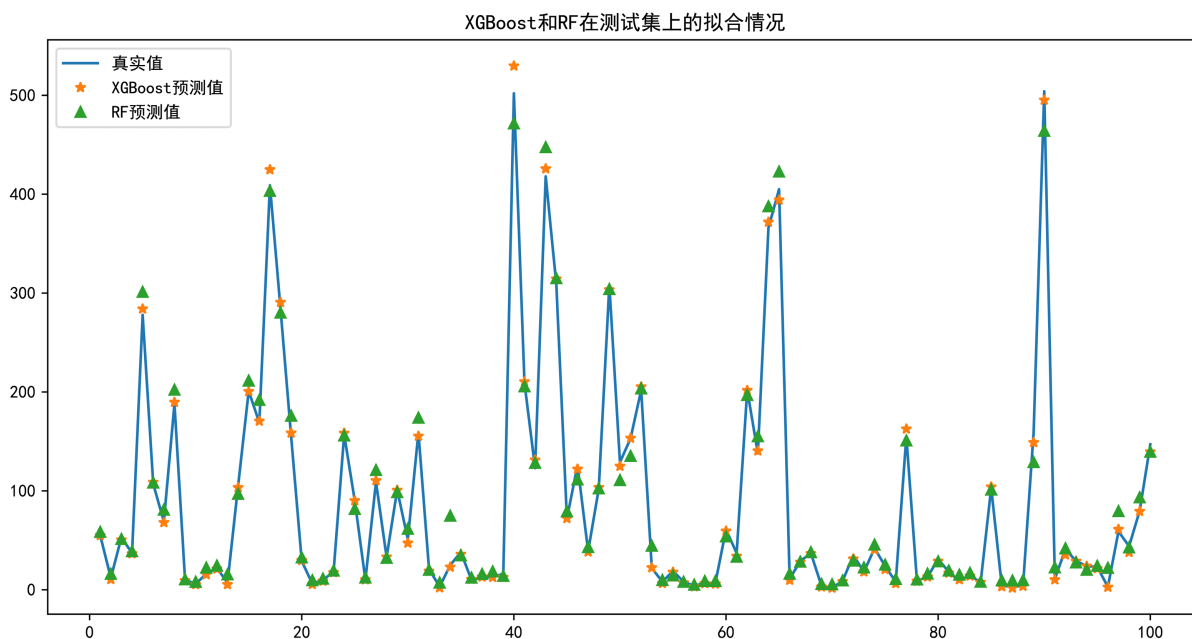


Figure 2. Fitting of XGBoost and RF on the test set
图 2. XGBoost 和 RF 在测试集上的拟合情况

利用评估损失函数对两种模型进行性能比较，采用均方误差(RMSE)，平均绝对误差(MAE)和对称平

均绝对百分比误差(SMAPE)作为预测模型的评价指标,误差越小说明拟合效果越好,模型准确率越高。评价指标如表 3 所示, XGBoost 模型的 $R^2 = 0.956$, RMSE = 36.522, MAE = 5.374, SMAPE = 16.718, 而 RF 模型的 $R^2 = 0.935$, RMSE = 44.303, MAE = 13.166, SMAPE = 37.531, XGBoost 模型的 R^2 、RMSE、MAE、SMAPE 均优于 RF 随机森林模型,故选用 XGBoost 模型进一步实现预测工程。

Table 3. Prediction error of two models

表 3. 两种模型的预测误差

评价指标	RF 随机森林	XGBoost
R^2	0.935	0.956
RMSE	44.303	36.522
MAE	13.166	5.347
SMAPE	37.531	16.718

4.4. XGBoost 模型预测

将预处理后的产品订单数据集以天为时间粒度划分为训练集与测试集两部分,利用 Python 构建预测模型,调整模型参数达到最优后,利用 XGBoost 模型对国内某大型制造企业未来 3 个月(即 2019 年 1 月、2 月、3 月)的月需求量进行预测,预测结果如表 4 所示。

Table 4. Prediction results

表 4. 预测结果

销售区域编码	产品编码	产品大类编码	产品细类编码	2019年1月预测需求量	2019年2月预测需求量	2019年3月预测需求量
101	20002	303	406	17.820	6.289	25.265
101	20003	301	405	643.455	224.751	164.411
101	20006	307	403	67.684	32.174	132.033
101	20011	303	401	-63.949	7.719	11.009
101	20014	307	403	2260.127	1383.639	1775.286
...
105	22066	307	403	1231.869	494.216	985.767
105	22072	305	412	-184.197	96.442	-507.315
105	22075	307	403	30.213	164.522	440.600
105	22083	303	401	611.825	383.923	1620.891
105	22084	302	408	-54.590	95.997	168.248

4.5. 小结

通过对原始数据进行构建特征工程,得到 20 个特征指标,并将其作为模型输入进行模型训练,再对训练好的模型进行网格结合手动调参优化后得到一个最优模型后,与 RF 进行对比,发现优化后的

XGBoost 模型的均方误差、平均绝对误差、对称平均绝对百分比误差均小于 RF, XGBoost 模型在订单需求量的预测表现出较好的性能。最后实现对企业未来 3 个月需求量的预测。

预测结果显示, 101 地区中产品类别组合为 301-405、303-401、303-406、305-412、306-402、307-403、308-404 的产品订单需求量总体呈增长趋势, 组合为 302-408、303-410、303-411、306-407 的产品订单需求量大部分在 1 月有降低趋势, 但在 2 月和 3 月又基本呈增长趋势, 组合为 304-409 的产品订单需求量大部分呈下降趋势。

102 地区中产品类别组合为 301-405、302-408、303-411、305-412、306-402、306-407、307-403、308-404 的产品订单需求量总体呈增长趋势, 组合为 303-401、303-410、304-409 的产品订单需求量大部分在 1 月有降低趋势, 但在 2 月和 3 月又基本呈增长趋势, 组合为 303-406 的产品订单需求量大部分呈下降趋势。

103 地区中产品类别组合为 301-405、302-408、303-401、303-411、305-412、306-402、307-403、308-404 的产品订单需求量总体呈增长趋势, 组合为 306-407 的产品订单需求量大部分在 1 月有降低趋势, 但在 2 月和 3 月又基本呈增长趋势, 组合为 303-406、303-410、304-409 的产品订单需求量大部分呈下降趋势。

105 地区中产品类别组合为 303-406、303-410、304-409、306-402、307-403 的产品订单需求量总体呈增长趋势, 组合为 306-407 的产品订单需求量大部分在 1 月有降低趋势, 但在 2 月和 3 月又基本呈增长趋势, 组合为 301-405、302-408、303-401、303-411、305-412、308-404 的产品订单需求量大部分呈下降趋势。

总体上看, 每个地区的订单需求量均有不同程度的波动, 其中 101 和 102 地区的订单需求量总体呈增长趋势, 而 105 地区的订单需求量下降程度较为严重。因此企业需要找出这些产品存在的问题, 并针对其做出相应举措, 进而提升该产品的订单需求量。

5. 结论

1) 将传统时间序列模型与集成学习方法对比分析, 讨论了 XGBoost 对于本文实现需求预测的优势与适用性, 通过优化提升了预测结果的准确性。并且选用 R^2 、RMSE、MAE、SMAPE 四项评估指标进行模型评估, 结果表明平均绝对误差为 5.347, 拟合优度高达 0.956, 说明了模型拟合度和精度都较为理想。同时, 本文还引入 RF 随机森林模型进行对比分析, 结果表明 XGBoost 的效果最优, 因此选择该模型进行预测是合理的。

2) 通过预测结果可知, 节假日期间订单量客观上相较于平时有明显涨幅, 表明人们在这个时期有更多的时间和精力来进行消费和购物, 因此商家可以利用这个时期提高产品和服务的销售额。此外, 商家可以根据节日需求做出相应调整来迎合消费者的需求。

3) 促销活动对提高销售有正向影响, 表明了消费者在购买产品时会被价格、优惠、礼品等促销手段所吸引, 从而产生购买欲望, 并增加了其对该产品的信任度。同时, 促销活动也是企业营销战略的一部分, 可以帮助企业快速推广和宣传品牌和产品, 提升知名度。

4) 春季和冬季的订单需求量较高, 表明存在季节性需求: 某些产品或服务在特定季节会有更高的需求, 比如冬季需要大量暖气或雪地工具, 春季需要园艺用品或户外运动装备等等。以及节日和假期需求: 春季和冬季都有一些重要的节日和假期, 比如春节、情人节、复活节、万圣节、圣诞节等等, 这些节日和假期会带来特定的消费需求。

基金项目

项目来源: 福建省科技厅, 项目名称: 基于 ATOT 技术的智能养老系统设计与开发, 项目编号: 2023350104000282。

参考文献

- [1] Kim, W.J. (2018) Overseas Construction Order Forecasting Using Time Series Model. *Korean Journal of Construction Engineering and Management*, **19**, 107-116.
- [2] Khashei, M. and Bijari, M. (2010) An Artificial Neural Network (p, d, q) Model for Timeseries Forecasting. *Expert Systems with Applications*, **37**, 479-489. <https://doi.org/10.1016/j.eswa.2009.05.044>
- [3] Voronin, S. and Partanen, J. (2014) Forecasting Electricity Price and Demand Using a Hybrid Approach Based on Wavelet Transform, ARIMA and Neural Networks. *International Journal of Energy Research*, **38**, 626-637. <https://doi.org/10.1002/er.3067>
- [4] 王斌, 杨抒, 贾清, 赵毅, 王业. ARIMA 模型在电商平台新疆红枣订单预测中的应用研究[J]. 福建电脑, 2019, 35(11): 5-8. <https://doi.org/10.16707/j.cnki.fjpc.2019.11.002>
- [5] 谭祖健, 官丹萍, 莫愁, 丁振伟, 李晓霞. 基于非平稳时间序列分析的发动机订单预测模型[J]. 中国新技术新产品, 2022(4): 16-19. <https://doi.org/10.13612/j.cnki.cntp.2022.04.005>
- [6] 施佳. 基于 ARIMA-BP 组合模型的某餐饮 O2O 企业订单预测研究[D]: [硕士学位论文]. 北京: 北京交通大学, 2018.
- [7] 欧红森, 姚玉南, 冯皓, 何溢. 基于 GA-灰色 BP 神经网络船舶备件需求预测模型[J]. 中国修船, 2023, 36(2): 41-46. <https://doi.org/10.13352/j.issn.1001-8328.2023.02.012>
- [8] 李秀丽, 孙国华. 基于机器学习的产品需求量预测方法研究[J]. 科技视界, 2020, 24(2): 98-99.
- [9] Zhao, Y., Qi, Y. and Dong, J. (2019) Product Demand Forecasting Based on Online Customer Reviews: A deep Learning Approach. *Journal of Business Research*, **96**, 365-377.
- [10] 刘建华, 李军. XGBoost 算法原理及应用[J]. 计算机科学, 2017, 44(12): 1-7.
- [11] 周志华, 吴恩达, 等. XGBoost: 基于提升树的高效机器学习算法[J]. 计算机学报, 2016, 39(3): 530-543.
- [12] 曹雷, 尚维, 谢士尧, 王向. 基于 AGNN 舆情指数网络的价格指数预测研究[J]. 管理学报, 2023, 20(3): 411-421+431.