

基于ELM-GBDT组合模型的生鲜农产品 线上需求趋势预测研究

盛依勤

江南大学商学院, 江苏 无锡

收稿日期: 2025年3月17日; 录用日期: 2025年3月31日; 发布日期: 2025年5月9日

摘 要

在生鲜农产品销售渠道中, 由于生鲜农产品的特殊性, 需要对于市场信息有较强的感知能力以协调供应链, 促进生鲜农产品供应链的高效稳定运行。而供应链特别是上游在市场实时信息对接方面存在明显不足问题。为此, 本文借助面向全平台公开的评论数据, 以提高供应链信息传递的实时性与准确性, 利用ERNIE模型进行评论文本的情感分析, 获取用户情感特征集, 并依据现有数据进行特征工程, 保留提取6个特征后, 通过PSO优化获取权重组合ELM-GBDT预测模型进行需求趋势预测。研究表明, 与单一模型相比, 组合预测模型显著提高了对生鲜农产品销量需求趋势的预测准确度, 同时发现生鲜农产品需求趋势变化存在一个相对稳定的周期, 为生鲜农产品供应链采购决策、库存优化决策等提供有价值的参考。

关键词

ERNIE情感分析, 特征工程, 机器学习, 生鲜产品, 需求预测

Forecasting Online Demand Trends for Fresh Agricultural Products Using an ELM-GBDT Ensemble Model

Yiqin Sheng

School of Business, Jiangnan University, Wuxi Jiangsu

Received: Mar. 17th, 2025; accepted: Mar. 31st, 2025; published: May 9th, 2025

Abstract

In the sales channels of fresh agricultural products, due to the unique nature of these products, a

strong market information perception capability is required to coordinate the supply chain and promote its efficient and stable operation. However, significant deficiencies exist in the real-time integration of market information, particularly in the upstream segments of the supply chain. To address this issue, publicly available review data across all platforms was leveraged in this paper to enhance the real-time accuracy of information transmission within the supply chain. The ERNIE model was used for sentiment analysis of review texts. A set of user sentiment features was extracted, and feature engineering was conducted based on existing data. Six key features were retained. Subsequently, through Particle Swarm Optimization (PSO), a combination of weights was obtained to construct an ELM-GBDT ensemble prediction model for demand trend forecasting. The research results indicate that, compared to single models, the ensemble prediction model significantly improves the accuracy of forecasting demand trends for fresh agricultural products. Additionally, it is found that there is a relatively stable cycle in the demand trends for fresh agricultural products, which provides valuable insights for procurement decisions and inventory optimization in the supply chain.

Keywords

ERNIE Sentiment Analysis, Feature Engineering, Machine Learning, Fresh Products, Demand Forecasting

Copyright © 2025 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,国家对生鲜农产品供应链体系建设的重视逐步加深。2021年5月为切实落实《中共中央国务院关于全面推进乡村振兴加快农业农村现代化的意见》要求,发布《关于进一步加强农产品供应链体系建设的通知》,其中重点强调需要完善农产品零售网点建设,强化产销对接长效机制,支持生鲜电商等各类农产品流通企业拓宽各类农产品营销渠道。根据网经社电子商务研究中心联合数字零售台发布的《2023年上半年中国生鲜电商市场数据报告》显示,2023年生鲜电商交易规模达6427.6亿元,同比增长14.74%。与此同时,随着整体生鲜电商市场规模、销售渠道的扩展,传统的生鲜农产品预测体系已经无法应对当前复杂、广泛的市场情况。由于生鲜农产品的特殊性,供需不平衡已经成为常态。其产品具有周期性和地域性,导致生产和销售在时间和空间上分离,从而导致供需不平衡呈现出发散状态。借助生鲜电商平台的有效信息流通模式有利于确保生鲜农产品的稳定供应,减少缺货或库存积压等损失,提升供应链效率。

信息共享在供应链管理中扮演着至关重要的角色,成为当前研究的核心焦点。先前的研究已经通过构建二级或多级供应链来阐述信息共享策略对牛鞭效应的影响[1]-[3],明确信息共享程度与总成本之间的负相关关系[4]。虽然信息共享对企业成本的降低所带来的益处,但需求信息共享的准确性及全面性会直接影响供应链成员的管理决策,未必能够实现供应链的正向互动。因此,完善信息的完整度与准确度显得尤为重要,特别是在数字化和网络化背景下,信息共享的实时性和动态性成为新的挑战。随着网购渠道的成熟,学者们逐渐将线上平台呈现的各种形式和内容视为新的信息源,以探索在线信息对产品销量、产品需求等之间的关系与影响。Wang (2022)通过研究在线评论的可信度感知与产品销量的关系,发现随着对在线评论可信度认知的提高,评论情感与产品销量之间的关系发生了变化[5]。另外,丁沛等(2022)应用在线评论中的评论数量、情感以及点赞等特征,以129款新能源车为研究样本,发现这些评论

特征对车辆销量产生了积极影响[6]。Zhang 等(2021)通过分析在线评论中关于竞争对手产品销量、产品价格、评论情感价值和产品网络热度的讨论,构建了销量预测模型[7]。杨亚琛等(2024)提出一种基于在线产品评论的竞争情报挖掘方法,将网络评论文本中的关键信息应用于乘用车需求趋势预测[8]。上述研究使用的训练方法大多为 Word2Vec 等传统方法,这些方法虽然在文本表示学习方面取得了一定成效,但在处理中文文本时,仍存在语义理解不深、上下文关联性不足等问题。近年来,随着深度学习技术的快速发展,基于 Transformer 的预训练语言模型(如 BERT、ERNIE 等)在文本表示学习中展现出显著优势。特别是 ERNIE (Enhanced Representation through Knowledge Integration),通过引入知识增强和多层次语义理解,在中文文本处理任务中表现尤为突出[9]。

在面对需求预测问题时,研究者们关注数据处理和模型优化的不同方面。Miguéis 等(2022)强调了销量预测与需求预测之间的差异,并提出了一种基于预测的处理数据缺失的数据审查方式,以减少行业末端浪费[10]。此外,陆佳圆等(2011)提出了基于变精度粗糙集和最小二乘支持向量机回归理论的集成方法,通过历史、当前和预测的权重数据为产品设计决策提供有力支持[11]。沈超等利用 Mann-Kendall 时间序列预测模型预测关注度和情感度波动,以进一步预测消费者需求的发展趋势[12]。徐小峰等(2022)在微观层面综合考虑时间、定价等因素进行特征分析,利用 NARX 弥补 ARIMA 模型在非线性关系捕捉上的不足,以提高生鲜产品短期销量的预测精度[13]。高晓辉等(2023)基于经验模态分解使用 XGBOOST 和 ELM 对不同特征分量进行分析,预测南京市空气质量并进行预警研究[14]。范亚茹等(2024)基于 VMD-CNN-GRU-LSTM 组合模型的汽车销量预测分析,针对复杂特征提出融合多种神经网络的方法[15]。王阳等(2024)提出一种基于订单时序和订单频率的改进自注意力机制模型对快消品的销量进行预测[16]。研究者们通过引入时间序列分析、非线性关系捕捉以及多特征融合等技术,不断提升预测精度。然而,尽管这些方法在各自领域取得了显著成果,但在面对复杂的中文语境和在线数据时,存在语义理解不足等问题。

当前研究展示了在线评论对产品需求的潜在影响,但是由于中文语境语义的特殊性,使用何种情感分析器来更为准确地提取在线文本中的情感信息,并以此作为特征投入需求预测的研究并不多。因此,本文将采用更适合中文语义分析的 ERNIE 提取文本中的情感信息做进一步研究。此外,一般单一模型的预测效果并不如组合模型,目前,针对线上生鲜产品需求趋势的预测研究中,使用组合模型的方法尚不多见。因此,本文尝试采用组合模型来预测线上生鲜产品的需求趋势,基于在线评论中蕴含的用户情感等相关因素,通过结合多个数据源和预测模型,建立包含多个特征因素的组合预测模型,以期对产品需求量进行准确的预测,从而使企业准确了解顾客需求,使得供应链中的需求信息实现及时共享。

2. 研究方法

本节将介绍生鲜产品需求趋势预测模型,重点介绍数据获取及文本数据处理方法,使用情感分析对生鲜产品评论数据构建情感特征集,以及如何结合特征工程处理后的特征使用 ELM-GBDT 组合模型对生鲜产品需求趋势进行预测。

2.1. 消费者情感指标构建

研究使用的预训练语言模型为 ERNIE 模型,与表现较好的 BERT 模型相比,改进掩码机制以适应中文语言特点,使模型能够训练学习到比单字更大的语义单元,更加适用于中文文本处理。其情感分类基本原理包括预训练阶段在大规模文本语料上学习语言知识,采用双向 Transformer 编码器提高语义表示能力,引入知识融合机制增强模型的语义信息,通过改进的掩码机制训练(图 1)学到更大的语义单元,最后微调适应情感分类任务。

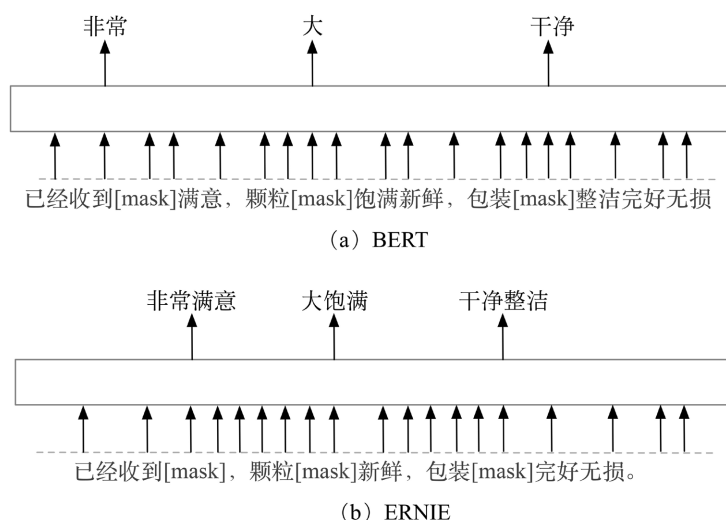


Figure 1. Different masking strategies for BERT and ERNIE

图 1. BERT 和 ERNIE 不同的掩码策略

基于对每条评论文本的情感分类(0, 1)计算结果构建情感指标, 构建过程如下:

Step1: 由于一天之内存在多条评论以及其对应的情感分类情况, 所以需要根据时间戳找到时间对应的情感分类结果, 按照天数对情感分类结果进行时间分组, 将每个日期下的情感分类结果生成一个列表, 汇总得到 *senti_classify_dataset*。

Step2: 统计每日列表中正向情感数量 *pos_num* 和负向情感数量 *neg_sum*, 以及两者相对比值, 构建方式如下:

$$senti_score = \frac{pos_num - neg_sum}{pos_num + neg_sum}$$

最终获得消费者情感特征集包含三个指标: *pos_num*、*neg_sum*、*senti_score*。其中, *pos_num*、*neg_sum* 属于绝对指标, *senti_score* 属于相对指标。

情感分类器的性能评价通常涉及多个指标, 一般可使用准确率(Accuracy)、精确度(Precision)、召回率(Recall)、F1 分数(F1 Score)等, 具体公式如下。

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

$$Precision = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

$$F1\ Score = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

其中, *TP* (True Positives)是正确预测为正类的样本数, *TN* (True Negatives)是正确预测为负类的样本数, *FP* (False Positives)是错误预测为正类的样本数, *FN* (False Negatives)是错误预测为负类的样本数。

2.2. 特征工程

特征工程是将原始数据转换为更能代表数据内涵的特征矩阵过程, 利用特征矩阵训练模型, 能够降

低模型计算成本、提升预测潜力。特征工程包括特征提取、特征衍生、特征选择及特征降维等步骤。本文研究背景下重点为特征衍生和特征选择步骤，介绍如下。

2.2.1. 特征衍生

在预测问题中，由于数据本身特征数量十分有限，因此需要构造一定的衍生特征来增强模型的拟合能力。基于历史数据，通过滑动窗口构建一系列统计特征、比值特征以及差分特征等(如图 2)。

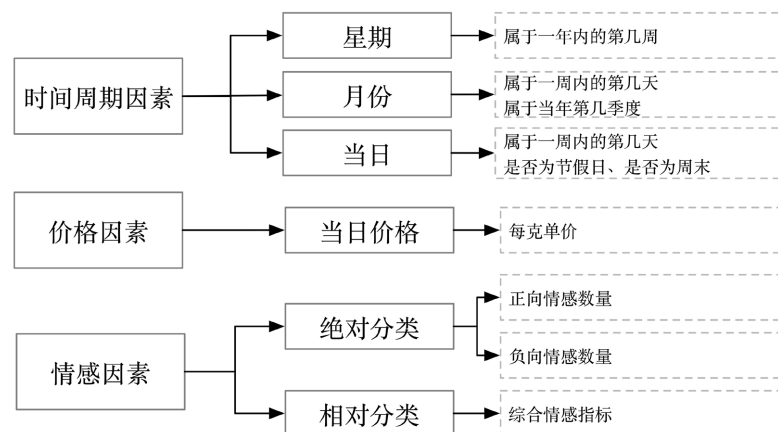


Figure 2. Description of feature indicators and feature derivation

图 2. 特征指标及特征衍生说明

2.2.2. 特征选择

本研究将选用随机森林进行特征的选择选取一定重要性阈值内的特征投入模型。随机森林的特征选择可以简化正在建模的问题，加快建模过程，一般可以提高模型的性能，并且提高模型的泛化能力。

在随机森林中，特征重要性的计算通常使用基尼不纯度等指标来衡量，以反映特征在决策树节点分裂中的贡献程度，基尼不纯度公式为

$$I(t) = \sum_{i=1}^C p(i|t) \cdot (1 - p(i|t)) \quad (1)$$

其中， $I(t)$ 为节点 t 的基尼不纯度， C 为类别数， $p(i|t)$ 为节点 t 中属于类别 i 的样本比例。

每棵决策树都会计算特征的重要性，然后通过对所有决策树的重要性进行平均来得到最终的特征重要性，以反映特征对于模型预测的贡献程度，分数越高表示对模型的影响越大，特征重要性公式为

$$Importance(j) = \frac{1}{N} \sum_{t=1}^N I(t, j) \quad (2)$$

其中， $Importance(j)$ 为表示特征 j 的重要性， N 是决策树的总节点数， $I(t, j)$ 表示在节点 t 上使用特征 j 进行分裂后基尼不纯度减少量。

2.3. ELM-GBDT 生鲜产品需求趋势预测模型

2.3.1. ELM 生鲜产品需求趋势预测

Huang 等人在 2004 年提出了具有快速学习速度的单个隐层前馈神经网络(ELM)。该方法不需要迭代调优才能达到全局最小值[17]。它已被证明具有良好的泛化性能、通用逼近能力[18]、高学习精度和最少的人工干预[19]。隐藏层节点的权重和偏差是随机设置的，从而节省了通过反向传播进行微调所需的时间。输出层的权重可以通过 Moore Penrose 逆运算解析确定。一个通用的 ELM 模型如图 3 所示。

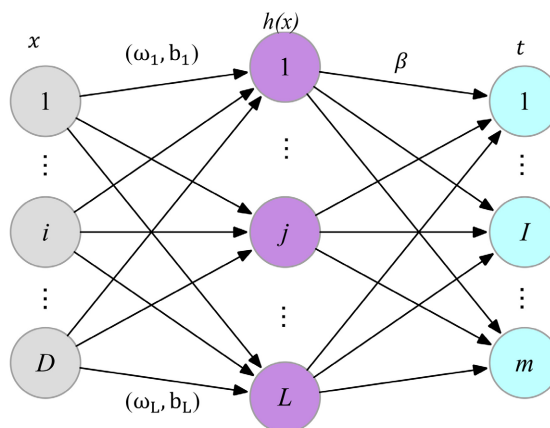


Figure 3. ELM extreme learning machine network structure diagram

图 3. ELM 极限学习机网络结构图

超参数选择方面,如表 1,使用网格搜索对隐藏神经元数量、随机状态和正则化参数进行寻优,使用 5 折交叉验证并通过均方误差来评估模型的性能。

Table 1. Description of the ELM hyperparameters

表 1. ELM 超参数表示

超参数	表示	参数选择范围
激活函数	<i>activation</i>	<i>tanh, relu, huber, gelu, swish</i>
隐藏层节点数	<i>L</i>	[10~100]
随机状态	<i>random_state</i>	[10~100]
正则化参数	<i>alpha</i>	[0.001~1]

2.3.2. GBDT 生鲜产品需求趋势预测模型构造

梯度提升决策树(Gradient Boosting Decision Trees, GBDT)是一种基于决策树的集成学习方法,属于监督学习的回归与分类算法,可以有效地捕获复杂的非线性关系。GBDT 一般使用一个简单的弱学习器作为首个预测模型,将其输出作为初始预测值 y_0 开始进行迭代。本文使用的损失函数为均方误差,对于第 m 轮迭代,计算当前模型在训练集上的负梯度,使用梯度提升树作为弱学习器进行拟合即拟合当前模型的残差,通过将新生成的弱学习器的预测结果与之前的预测结果加权相加,来更新模型的预测值。考虑到使用的数据集量,以设定的迭代轮数为终止条件终止迭代。

超参数选择方面,如表 2,使用网格搜索对学习率、弱学习器数、每个模型使用的样本比例和每个模型最大深度进行寻优,使用 5 折交叉验证并通过均方误差来评估模型的性能。

Table 2. Hyperparameter representation in GBDT algorithm

表 2. GBDT 算法中超参数表示

超参数	表示	参数选择范围
学习率	γ , <i>learning_rate</i>	[0.01~0.9]
弱学习器数	<i>n_estimators</i>	[100~1000]
每个模型使用样本比例	<i>subsample</i>	(0~0.9]
每个模型最大深度	<i>sax_depth</i>	(2,3,4,5,6,7,8)

2.3.3. 组合方法

基于 ELM-GBDT 模型对生鲜产品需求趋势预测步骤, 如图 4 所示。

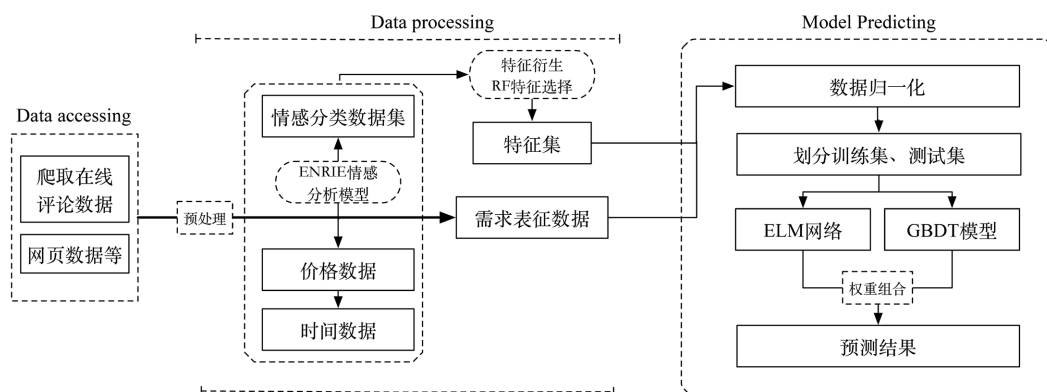


Figure 4. ELM-GBDT demand forecasting framework based on feature engineering

图 4. 基于特征工程的 ELM-GBDT 需求预测框架

在完成数据获取和情感分析过程后, 本研究采用了随机森林方法对时间特征数据和情感分类特征进行选择。对输入数据进行预处理, 将其转换为数组形式, 并进行 Z-score 归一化处理, 以确保数据的一致性和可比性。按照时间顺序将数据划分为训练集和测试集, 其比例为 8:2。在模型构建阶段, 采用了极限学习机(ELM)网络和梯度提升决策树(GBDT)模型进行训练和测试, 并对 ELM 和 GBDT 的超参数进行了选择和调优。

在模型预测阶段, 基于 GBDT 和 ELM 获得的预测结果, 将结果进行加权获得最终的预测结果。由于不同模型具有不同的性能和表现, 多个权重值设置能够平衡不同模型在组合中的贡献, GBDT 通过特征自动分箱和迭代提升, 擅长捕捉时间序列中的非线性交互和稀疏情感特征的局部模式, ELM 随机权重初始化与单隐层结构使其对数据尺度不敏感且适合快速处理高维时间特征, 加权结果可缓解 GBDT 对噪声过拟合的风险以及 ELM 对复杂模式欠拟合的问题, 因此, 为两个模型设置 2 个权重值 ω_1 和 ω_2 , 最终的组合模型预测结果记为 $Y(x)$:

$$Y(x) = \omega_1 \cdot G_L(x) + \omega_2 \cdot F(x)$$

其中 $G_L(x)$ 是 ELM 对应的输出函数即模型预测结果, $F(x)$ 是 GBDT 对应的预测结果。

为求解最优的权重组合, 使用 PSO 算法进行寻优求解, 以均方误差最小为目标, 采用粒子群算法对参数进行优化, 操作步骤如下:

- 1) 种群初始化。设置迭代次数、种群大小、学习因子、惯性权重等基本参数。
- 2) 计算该算法的适应度函数值, 并将 MSE 作为算法的适应度函数。
- 3) 根据适应度函数计算粒子个体及全局最优解, 并与历史值进行比较更新群体中粒子的速度和位置。
- 4) 判断是否满足终止条件, 如满足则停止迭代, 若不满足则不断更新粒子速度和位置直至满足终止条件。

2.3.4. 其他模型

本文使用的主要对比模型为 XGBoost。XGBoost (eXtreme Gradient Boosting) 是一种高效的梯度提升决策树算法, 由华盛顿大学的陈天奇博士提出, 并在 Kaggle 的希格斯子信号识别竞赛中使用。XGBoost 是基于梯度提升树(Gradient Boosting Decision Tree, GBDT)的算法, 属于集成学习算法中的一种, 通过组合多个弱学习器来形成一个强学习器。其核心思想是通过逐步优化损失函数来构建一个强学习器。每个弱学习器

是一个决策树模型，而这些决策树通过逐步优化损失函数来构建一个强学习器。在原有 GBDT 的基础上进行了改进，加入了正则化项以控制模型复杂度。虽然 XGBoost 基于 GBDT 进行改进，但是在某些特定任务中，GBDT 的表现可能优于 XGBoost。预测客户汽车保险索赔的研究中，GBDT 的 AUC 值略优于 XGBoost [20]。另一项研究中因为其参数较多且过拟合风险较高 XGBoost 并没有取得比 GBDT 更好的效果[21]。

3. 实证分析

3.1. 数据获取及处理

为验证评论情感指标对线上生鲜产品的需求预测准确度的影响，本文使用 python 爬虫代码获取生鲜农产品的评论数据，以京东超市的百香果为例，利用 python 爬取 2023 年 12 月 1 日~2024 年 2 月 8 日的评论数据，得到 4043 条数据，部分数据见表 3。根据网站的评论规则，每个顾客在确认收货后均会形成一条文本评论数据，基于此将每天的评论数量作为生鲜产品的需求表征数据，以天为预测周期进行生鲜产品需求趋势预测。

Table 3. Partially crawled review data
表 3. 部分爬取的评论数据

Date	Content
.....	
2023-12-01 09:25:00	百香果酸酸甜甜的，挺好吃的！
2023-12-01 10:14:29	这个百香果巨甜，味道很好很正，冲蜂蜜水也巨好喝。
2023-12-01 10:02:35	都吃完了没图片，但这百香果很香，味道不错，放久后皮虽然会变皱，但里面的果肉没有问题，和柠檬一起泡蜂蜜喝十分的美味，推荐大家品尝！
2023-12-01 12:55:33	纯酸。收到青的。放着变皱了。但还是青。吃着纯酸。还是黄色好吃。
2023-12-01 15:02:44	百香果收到了，很新鲜，个头很大，个头大，个头均匀，很甜很好吃，味道酸酸甜甜真不戳，快递也很快，包装没有损坏，超爱吃嘿嘿。
.....	

3.2. 情感特征集构建

为了验证 ERNIE 模型的准确性，本文选择 BERT 和 SVM 情感分析作为对比模型，从各项性能看，本文使用的 ERNIE 模型在准确率、精准度、召回率和 F1 分数的表现均达到 91%以上，其整体分类效果更优，具体情况见表 4。

Table 4. Sentiment classifier performance comparison
表 4. 情感分类器性能对比

性能	ERNIE	BERT	SVM
Accuracy	93.70%	91.28%	81.78%
Precision	95.52%	87.31%	92.94%
Recall	91.96%	96.21%	68.78%
F1 score	93.71%	91.55%	79.04%

因此使用 ERNIE 对评论情感进行分析具有合理性，ERNIE 实例分类情况如表 5。通过 ERNIE 进行情感分析后的部分统计结果如表 6，根据情感指标计算公式获得每日情感指标值 *senti_score*，由 *neg_sum*、*pos_sum* 和 *senti_score* 三个指标共同构成情感特征集。

Table 5. ERNIE instance classification results
表 5. ERNIE 实例分类情况

Date	Comments	Result
2022-04-30	百香果的质量非常好，一直在用的，也真的很喜欢，完全超出期望值。发货速度非常快，包装非常仔细、严实，物流公司服务态度很好，运送速度很快。很满意的一次购物，包装整洁干净，希望店家多多优惠，及时通知老客户，促成回购。祝生意兴隆。	1
2023-12-22	真的真的特别特别酸。	0
2022-03-28	物美价廉，很 OK，比实体店优惠太多了物流速度真的是没话说，就是一个字：快。商品从外包装来看非常高档，真是一分钱一分货，贵有贵的道理。	1
2022-03-06	见货非常满意，颗粒大饱满新鲜，包装干净整洁完好无损。	1
.....		

Table 6. Partial statistical results after sentiment analysis
表 6. 情感分析后的部分统计结果表

Date	统计	neg_sum	pos_sum
2023-12-01 00:00:00	[1, 1, 1, 1, 1, 1, 0, 0, 1, 0, 0, 0, 0, 1, 0]	7	8
2023-12-02 00:00:00	[1, 1, 1, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0]	12	5
2023-12-03 00:00:00	[1, 1, 1, 1, 1, 1, 1, 1, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 1, 0, 0, 0, 0]	18	10
2023-12-04 00:00:00	[1, 1, 1, 1, 0, 1, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0]	10	7
2023-12-05 00:00:00	[1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 1, 0, 0, 0]	13	4
.....			

3.3. 特征工程

基于上述数据处理情感特征集构建，展开进一步的特征衍生和特征选择处理，尽可能挖掘有效信息，以获得高质量的输入特征数据，提高对于生鲜产品需求趋势预测的准确性。

3.3.1. 特征衍生

考虑到特征数据维度的有限性，在已有数据集的基础上从时间和情感两因素进行扩充，构建 8 个时间因素特征、3 个情感因素特征和 1 个价格因素特征，共 12 个初步特征，见表 7，其中真正对预测结果有帮助的特征需要做进一步筛选选择。

Table 7. Feature sets after derivation
表 7. 衍生后的特征集

有关因素	说明	变量表达
时间因素	属于一年内的第几周	<i>week_of_year</i>
	属于一周内的第几天	<i>day_of_week</i>
	属于当年的第几季度	<i>quarter</i>
	是否为节假日	<i>if_holiday</i>
	是否是周末	<i>if_weekend</i>
	是否为节假日前 3 天	<i>holiday_before_3d</i>
	是否为节假日前 1 天	<i>holiday_before_1d</i>
	是否为节假日后 1 天	<i>holida_t_after_1d</i>

续表

情感因素	正向情感数量	<i>pos_num</i>
	负向情感数量	<i>neg_num</i>
	综合情感指标	<i>senti_score</i>
价格因素	当日单价	<i>price</i>

3.3.2. 特征选择

采用随机森林对上述 12 个特征进行特征选择(如图 5), 根据研究问题, 设置重要性阈值为所有特征重要性均值, 即 0.069, 选择除目标值外高于此阈值的特征, 分别为 *senti*、*neg_num*、*pos_num*、*weekofyear*、*dayofweek*、*price* 共六个特征。

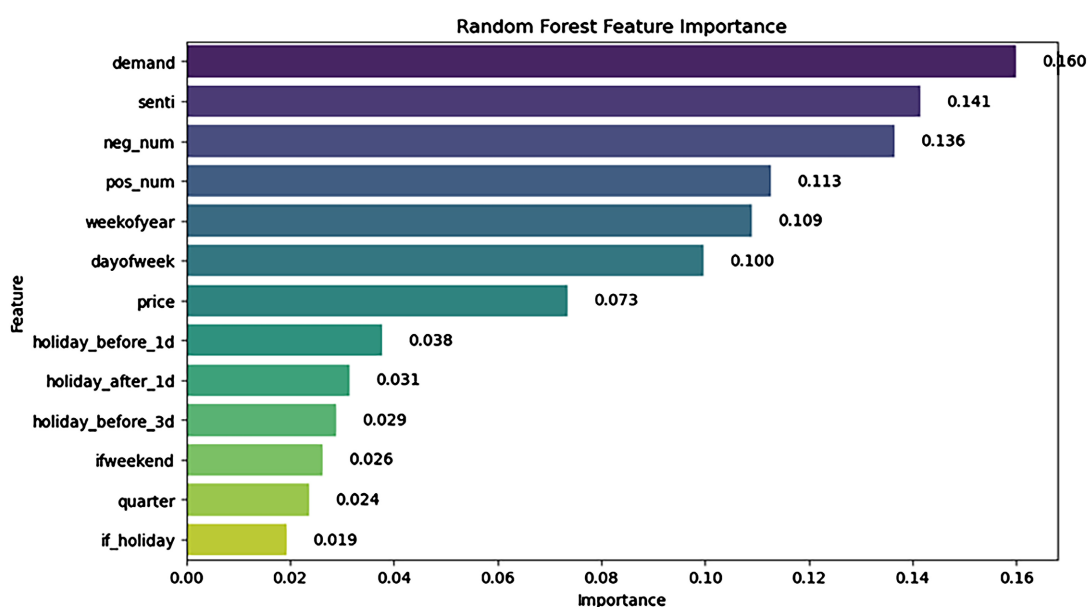


Figure 5. Random forest importance feature selection

图 5. 随机森林重要性特征选择图

3.4. 模型结果与分析

经过特征选择后得到 6 个输入特征, 分别是综合情感指标、正向情感数、负向情感数、当周属于一年中的第几周、当天属于一周中的第几天以及当日单价。

在此基础上对数据进行标准化处理, 将 2023 年 12 月 1 日~2024 年 2 月 10 日共 72 天的数据作为训练集, 分别用于训练 ELM、GBDT 模型, 求解最优参数, 将 2024 年 2 月 11 日~2024 年 2 月 28 日的数

据作为测试集, 用于检验模型。
其中, ELM 模型训练中确定最优隐藏神经元数量为 22, 最优随机状态为 30, 最优正则参数为 0.1。GBDT 模型训练中超参数选择, 学习率为 0.1, 选择的最优迭代数为 150, 每个弱学习器的最大深度为 2, 每个模型训练时使用的样本比例为 0.8, 终止条件为迭代次数。XGBoost 训练中超参数选择, 学习器数为 900, 列采样比例为 1, 学习率为 0.05, 最大深度为 2, 样本采样比例为 0.6。

经过 PSO 寻优, 以均方误差为目标函数, 设置迭代次数为 500, 种群大小为 50, 学习因子为 0.5, 惯性权重为 0.1, 求解得到, ELM 的最优权重为 0.43, GBDT 的最优权重为 0.57, 最后获得 ELM-GBDT

组合模型预测结果,并与 XGBoost 进行对比,在训练样本中得到的单个模型和组合模型预测值对比见图 6,经过前期数据构建与处理后的 ELM、GBDT 单一模型结果的波动较为明显,ELM-GBDT 组合模型的预测结果波动较小,整体更接近真实需求趋势。

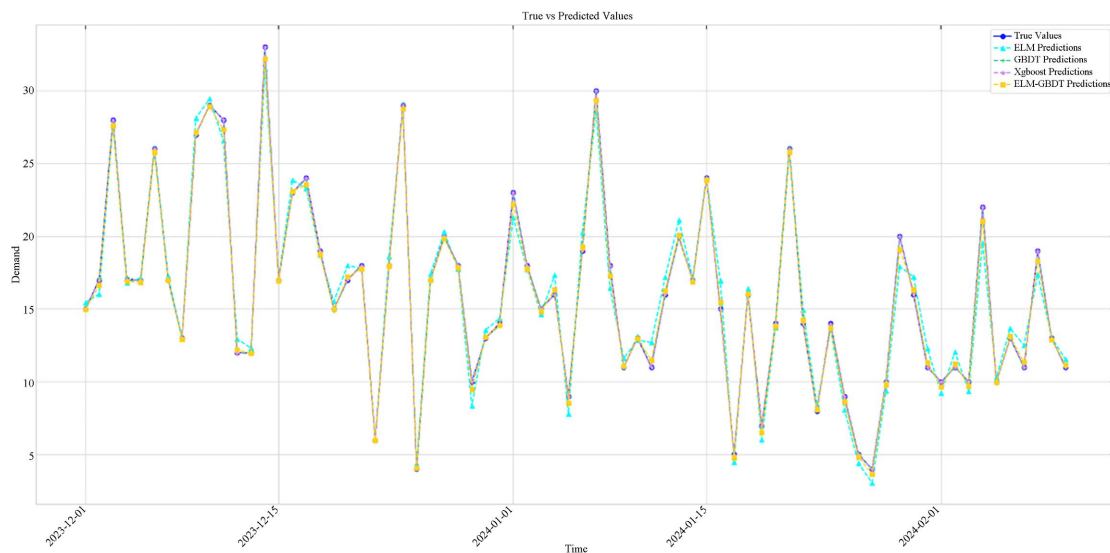


Figure 6. Comparison of fit between single model and combined model in training samples

图 6. 训练样本中单一模型和组合模型拟合对比

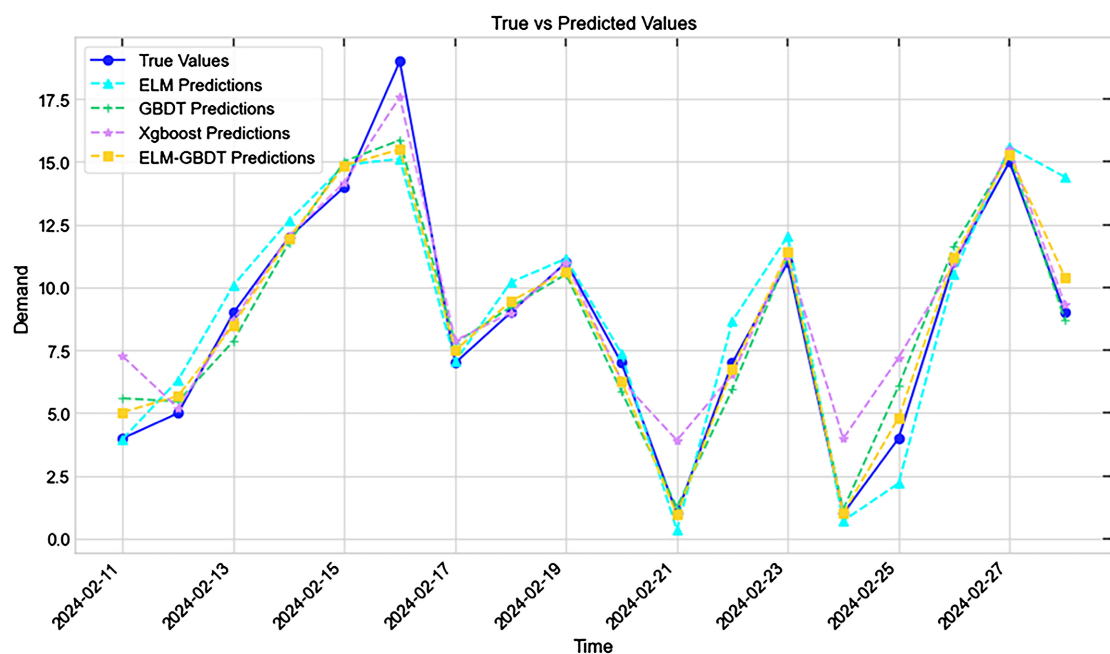


Figure 7. Comparison of fit between single model and combined model in training samples

图 7. 测试样本中单一模型和组合模型拟合对比

利用建立好的组合模型对测试样本的百香果需求数据进行预测,结果见图 7。

为验证三种模型对于需求趋势预测的准确性,分别计算三种模型的均方根误差 RMSE、平均绝对百分比误差 MAPE 和判定系数 R^2 。其中 RMSE 越小代表预测性能越好,MAPE 越接近 0 表示模型预测精

度越好, R^2 在 0~1 之间, 越接近 1 表示模型对实际数据的拟合程度越好。具体计算结果如表 8 所示, 发现 ELM-GBDT 的预测误差均优于单一模型, 预测精度较高。因此, ELM-GBDT 模型对于生鲜农产品需求趋势具有更好的预测能力。

Table 8. Comparison table of model prediction results evaluation

表 8. 模型预测结果评估对比表

NO.	MODEL	RMSE	MAPE(%)	R^2
1	ELM	1.7967	18.5290	0.8494
2	GBDT	1.1303	10.3328	0.9404
3	XGBoost	1.5351	19.85683	0.8900
4	ELM-GBDT	0.9128	6.5738	0.9611

基于上述模型拟合评估结果, 使用训练好的模型对未来 15 天生鲜农产品需求情况进行预测, 见图 8。

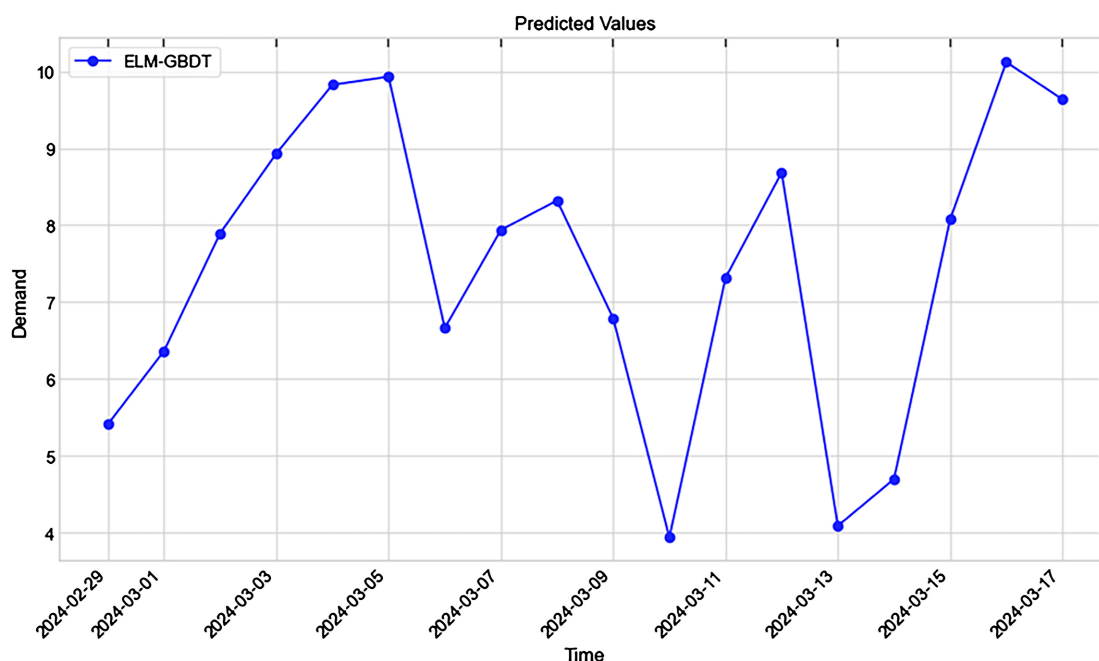


Figure 8. Results of passion fruit demand forecast for the next 15 days

图 8. 未来 15 天百香果需求预测图

在短期内, 效率需求呈现上升趋势, 在未来 5 天达到需求较高点, 随后的需求存在一定波动, 在未来第 10 天达到需求较低点, 在未来第 14 天需求达到又一高点, 存在一定波动。针对生鲜农产品高需求和低需求出现频率较高的情况, 线上生鲜供应链企业应该建立紧密的供应链合作关系, 确保及时了解产品供应情况。高需求趋势期间, 要合理增加库存水平以满足订单需求, 但同时密切监控库存周转率, 提高库存周转率至每周 3 次, 避免过度库存造成浪费。低需求或者波动需求趋势期间, 根据需求预测和销售趋势, 调整采购计划, 适度减少订购量, 并与供应商协商调整交付时间以灵活应对订单量和交货周期, 避免过早下单或过量采购, 以减少库存积压和资金损失。从特征指标看, 高需求趋势期间, 需关注消费者评论中所呈现的情感信息, 注重对于客户的服务水平, 设定客户服务响应时间, 及时处理订单变更、退换货等问题, 减少需求压力下的差评率, 在低需求或者波动需求趋势期间, 在价格等制定方面可实行

优惠策略,可以根据一定周期调整价格计划,以刺激生鲜农产品需求。对于并不直接接触线上市场销量情况的生鲜农产品供应商来说,可根据需求预测情况,阶段性调整不同渠道分配比重,及时调整生鲜产品品种和供应量。

综上所述,根据本研究的研究结果,企业可结合自有数据在库存管理、客户服务、营销策略等方面进行优化调整,有助于提高供应链效率,减少库存积压,提升客户满意度,从而在竞争激烈的市场中获得优势。本研究为解决现代企业管理中信息共享问题提供了新的视角和方法。

4. 结论

本研究以生鲜农产品为研究对象,结合语言模型以及应用机器学习模型研究其未来需求趋势。通过更适合中文文本分析的 ERNIE 模型进行文本情感的分类,构建综合情感指标,并构建 ELM-GBDT 预测模型,在现有数据集基础上提取 6 个特征,提高了对于生鲜农产品线上需求趋势预测的准确性和精度并使用模型对未来 15 天的生鲜产品需求情况进行预测,辅助商家和供应商做出更科学的库存管理和营销决策等,为应对市场短期波动提供了策略调整支持。

尽管当前研究在数据多样性、模型泛化能力、实时性、用户行为分析、供应链协同以及政策市场环境考量等方面存在局限,未来的拓展将整合更广泛的数据源,强化模型的实时处理和泛化应用,深入分析用户行为,优化供应链流程,并考虑外部环境变化,以期为生鲜电商行业的可持续发展贡献更多价值。

基金资助

江苏省研究生科研与实践创新计划项目(SJCX23_1221)。

参考文献

- [1] Cachon, G.P. and Fisher, M. (2000) Supply Chain Inventory Management and the Value of Shared Information. *Management Science*, **46**, 1032-1048. <https://doi.org/10.1287/mnsc.46.8.1032.12029>
- [2] Chen, F., Drezner, Z., Ryan, J.K. and Simchi-Levi, D. (2000) Quantifying the Bullwhip Effect in a Simple Supply Chain: The Impact of Forecasting, Lead Times, and Information. *Management Science*, **46**, 436-443. <https://doi.org/10.1287/mnsc.46.3.436.12069>
- [3] Dejonckheere, J., Disney, S.M., Lambrecht, M.R. and Towill, D.R. (2004) The Impact of Information Enrichment on the Bullwhip Effect in Supply Chains: A Control Engineering Perspective. *European Journal of Operational Research*, **153**, 727-750. [https://doi.org/10.1016/s0377-2217\(02\)00808-1](https://doi.org/10.1016/s0377-2217(02)00808-1)
- [4] Lin, F.-R., Huang, S. and Lin, S.-C. (2002) Effects of Information Sharing on Supply Chain Performance in Electronic Commerce. *IEEE Transactions on Engineering Management*, **49**, 258-268. <https://doi.org/10.1109/tem.2002.803388>
- [5] Wang, Q., Zhang, W., Li, J., Mai, F. and Ma, Z. (2022) Effect of Online Review Sentiment on Product Sales: The Moderating Role of Review Credibility Perception. *Computers in Human Behavior*, **133**, Article ID: 107272. <https://doi.org/10.1016/j.chb.2022.107272>
- [6] 丁沛, 马铁驹, 马也. 基于在线评论的新能源汽车销量影响因素研究[J]. 系统科学与数学, 2022, 42(10): 23-48.
- [7] Zhang, G. and Qiu, H. (2021) Competitive Product Identification and Sales Forecast Based on Consumer Reviews. *Mathematical Problems in Engineering*, **2021**, Article ID: 2370692. <https://doi.org/10.1155/2021/2370692>
- [8] 杨亚琛, 李全森, 唐浩冬. 基于在线评论的乘用车需求趋势分析方法[J]. 运筹与管理, 2024, 33(1): 76-82.
- [9] 余同瑞, 金冉, 韩晓臻, 等. 自然语言处理预训练模型的研究综述[J]. 计算机工程与应用, 2020, 56(23): 12-22.
- [10] Miguéis, V.L., Pereira, A., Pereira, J. and Figueira, G. (2022) Reducing Fresh Fish Waste While Ensuring Availability: Demand Forecast Using Censored Data and Machine Learning. *Journal of Cleaner Production*, **359**, Article ID: 131852. <https://doi.org/10.1016/j.jclepro.2022.131852>
- [11] 陆佳圆, 冯毅雄, 谭建荣, 等. 产品顾客需求权重的动态趋势预测与分析[J]. 计算机集成制造系统, 2011, 17(10): 2129-2136.
- [12] 沈超, 王安宁, 方钊, 等. 基于在线评论数据的产品需求趋势挖掘[J]. 中国管理科学, 2021, 29(5): 211-220.
- [13] 徐小峰, 余乐安, 林姿汝, 等. 基于特征融合的生鲜商品短期销量组合预测[J]. 管理科学学报, 2022, 25(12): 102-123.

-
- [14] 高晓辉, 周坤, 李廉水. 基于 XGBOOST 和 ELM 的混合空气质量预警系统: 以南京为例[J]. 中国管理科学, 2023, 31(5): 269-278.
- [15] 范亚茹, 向兵. 基于 VMD-CNN-GRU-LSTM 组合模型的汽车销量预测分析[J]. 西南民族大学学报自然科学版, 2024, 50(4): 441-446.
- [16] 王阳, 何利力, 郑军红. 基于改进注意力机制 Transformer 网络的快消品销量预测方法[J]. 智能计算机与应用, 2024, 14(1): 175-179.
- [17] Huang, G.-B., Zhou, H., Ding, X. and Zhang, R. (2012) Extreme Learning Machine for Regression and Multiclass Classification. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, **42**, 513-529. <https://doi.org/10.1109/tsmcb.2011.2168604>
- [18] Huang, G., Chen, L. and Siew, C. (2006) Universal Approximation Using Incremental Constructive Feedforward Networks with Random Hidden Nodes. *IEEE Transactions on Neural Networks*, **17**, 879-892. <https://doi.org/10.1109/tnn.2006.875977>
- [19] Huang, G., Huang, G., Song, S. and You, K. (2015) Trends in Extreme Learning Machines: A Review. *Neural Networks*, **61**, 32-48. <https://doi.org/10.1016/j.neunet.2014.10.001>
- [20] Zheng, H.R., Peng, F., Tian, Y.W., Zhang, Z.Z. and Zhang, W.T. (2023) Insurance Fraud Detection Based on XGBoost. *Academic Journal of Computing & Information Science*, **6**, 68-74.
- [21] 闫宏飞, 胡扬, 王彦恺. 面向 Canvas 学习管理系统的在线数据分析和挖掘研究[J]. 文献与数据学报, 2024, 4(1): 72-85.