Artificial Intelligence and Robotics Research 人工智能与机器人研究, 2024, 13(3), 662-672
Published Online August 2024 in Hans. https://www.hanspub.org/journal/airr
https://doi.org/10.12677/airr.2024.133067

基于Python技术对农产品价格预测及系统开发

周晓娅[1*]，何松芝[2]

[1]重庆对外经贸学院，数学与计算机学院，重庆
[2]重庆对外经贸学院，大数据与智能工程学院，重庆

收稿日期：2024年7月5日；录用日期：2024年8月21日；发布日期：2024年8月30日

摘　要

随着信息技术和大数据的发展，农产品价格预测对市场分析和决策起着越来越重要。本文爬取了农产品从2022年1月1日至2024年6月23日的数据，并基于这些数据建立了线性回归模型，ARIMA模型，以及随机森林模型三种不同的模型进行预测。研究结果表明，将时间与地点作为自变量预测价格时，随机森林模型预测的效果优于其他两个模型，能有效地捕捉价格变化趋势，为市场参与者提供决策支持。本文基于随机森林模型，利用FLASK框架构建了WEB端，使用者只需要在该网页选择产地及时间便可直接看到当日所预测的价格。

关键词

时间序列，价格预测，机器学习，ARIMA模型，随机森林

Agricultural Products Price Prediction and System Development Based on Python Technology

Xiaoya Zhou[1*], Songzhi He[2]

[1]School of Mathematics and Computer Science, Chongqing College of International Business and Economics, Chongqing
[2]School of Big Data and Intelligent Engineering, Chongqing College of International Business and Economics, Chongqing

Received: Jul. 5th, 2024; accepted: Aug. 21st, 2024; published: Aug. 30th, 2024

Abstract

With the development of information technology and big data, agricultural product price forecasting

*通讯作者。

文章引用: 周晓娅, 何松芝. 基于 Python 技术对农产品价格预测及系统开发[J]. 人工智能与机器人研究, 2024, 13(3): 662-672. DOI: 10.12677/airr.2024.133067

plays an increasingly important role in market analysis and decision-making. This paper crawls the data of agricultural products from 1 January 2022 to 23 June 2024, and based on these data, three different models, linear regression model, ARIMA model, and random forest model, are established for prediction. The results of the study show that when time and place are used as independent variables to predict prices, the Random Forest Model predicts better than the other two models, effectively capturing price trends and providing decision support for market participants. Based on the Random Forest model, this paper constructs a web page using the FLASK framework, in which users only need to select the origin and time to see the predicted price on the same day directly.

Keywords

Time Series, Price Forecasting, Machine Learning, ARIMA Model, Random Forests

1. 引言

在全球化和市场经济背景下，农产品价格波动对经济体系产生深远影响。信息技术和大数据技术的发展为价格分析和预测提供了新的工具。价格的不稳定不仅影响农民的种植决策和收益，也关系到市场供需平衡和消费者的生活成本。随着信息技术的快速发展，大数据技术为农产品价格分析和预测提供了新的视角和工具。通过收集和分析历史价格数据，可以揭示价格变化的规律，为市场参与者提供决策支持。

现有的农产品价格研究主要集中在价格形成机制、影响因素分析以及价格波动特征等方面。刘峰[1]以白菜月价格数据为例，构建了 ARIMA 模型预测白菜未来的价格，研究表明 ARIMA (0, 1, 1)模型能够很好的模拟并预测白菜的价格。随着机器学习技术的发展，越来越多的学者开始尝试使用这些方法对价格进行预测，如支持向量机(SVM)、随机森林(RF)、神经网络等。张瑞瑜[2]基于舆情和深度学习预测算法，提出了改进的 LSTM 模型预测小宗农产品价格，研究结果表明该模型预测价格较好。郭锋[3]结合时间序列预测模型以及机器学习，构建了 ARIMA-SVR 模型对大蒜的价格进行预测，通过对比发现，该模型预测效果优于单一的 ARIMA 模型和 SVR 模型。马小菁[4]构建了传统的 ARIMA-EGARCH 模型，LSTM 神经网络以及基于 EEMD 的 LSTM-AR 模型预测山东大葱未来的价格。研究表明，基于 EEMD 的 LSTM-AR 模型最优，其次是传统的 ARIMA-EGARCH 模型，LSTM 神经网络的准确率最低。

然而，将多元统计分析与机器学习技术相结合，考虑多种因素对农产品价格的影响，进行综合预测的研究还相对较少。本文首先利用 Post 请求，Josn 接受技术访问与收集特定的数据。然后利用 Descrip 进行描述性统计分析，以掌握数据的基本特征，对异常值和缺失值进行处理。其次分析日期、产地、平均价、最低价、最高价对预测的影响，发现有平均价指标存在会导致过拟合，因此对平均价指标进行了剔除。此外，利用 ACF 和 PACF 图识别价格数据的内在相关性，选择合适的 ARIMA 模型参数进行时间序列分析。最后，选用随机森林模型和线性回归模型，通过交叉验证和参数调优，提高预测的准确性。通过对比分析，将日期和产地考虑为自变量，价格考虑为因变量，可以得到随机森林模型的精确度最高，其次是 ARIMA 模型，线性回归模型精度最差。在系统的实现上，本文采用 FLASK-WEB 框架，利用 Templates 文件建立 Index. Html 进行页面展示，Result. Html 进行数据处理，static 文件添加图片完善了 WEB 端。利用了 Css，Javascript，Html 技术简单实现了本系统的开发。

2. 理论基础

2.1. 线性回归模型

多元线性回归模型利用历史数据，可以考虑多种因素对一个因变量的影响。其本质是想办法找到自变量与因变量之间关系的数学表达式。设 y 为因变量，自变量为 $x_1, x_2, \cdots, x_k$，多元回归模型的一般形式为 $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k + \varepsilon$，其中 β_0 是回归常数，$\beta_1, \beta_2, \cdots, \beta_k$ 是回归系数，ε 是常数项。在 PYTHON 中利用 Statsmodels. api 和 Numpy 这两个第三方库进行实现，利用 Numpy 库的 df 方法(x = df [ ] 和 y=df[ ])，对自变量和因变量进行确定，使用 Statsmodels. api 中的 ols 进行拟合。

2.2. ARIMA 模型

C.P. Box 和 G.M. Jenkins [5]是时间序列分析的先驱，他们对发展自回归积分滑动平均(ARIMA)模型做出了关键贡献。这种统计方法在金融、经济和气象学等领域广泛用于分析和预测数据。ARIMA 模型的核心在于将非平稳序列转化为平稳状态，以便使用自回归(AR)和移动平均(MA)技术进行分析。模型中“p”和“q”分别代表自回归和移动平均的阶数。假设时间序列 $\{x_t\}$ 具有零均值，ARMA (p,q) 模型可表示为：

$$\begin{cases} x_t = \phi_0 + \phi_1 x_{t-1} + \cdots + \phi_p x_{t-p} + \varepsilon_t - \theta_t \varepsilon_{t-1} - \cdots - \theta_q \varepsilon_{t-q} \\ \phi_p \neq 0,\ \theta_q \neq 0 \\ E(\varepsilon_t) = 0,\ \mathrm{Var}(\varepsilon_t) = \sigma_\varepsilon^2,\ E(\varepsilon_t \varepsilon_s) = 0,\ s \neq t \\ E(x_s \varepsilon_t) = 0,\ \forall s < t \end{cases}$$

中心化 ARMA (p,q) 模型可以简写为

$$x_t = \phi_1 x_{t-1} + \cdots \phi_p x_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \cdots - \theta_q \varepsilon_{t-q}$$

其默认与 AR 模型、MA 模型条件相同。引进延迟算子，ARMA (p,q) 模型简记为

$$\Phi(B) x_t = \Theta(B) \varepsilon_t$$

式中，$\Phi(B) = 1 - \varphi_1 B - \cdots \varphi_p B^p$，为 p 阶自回归系数多项式，$\Theta(B) = 1 - \theta_1 B - \cdots \theta_q B^q$，为 q 阶移动平均系数多项式[6]。ARIMA (p,d,q) 模型中的“d”指差分次数，对于非平稳的序列，需要首先对序列进行平稳化处理，而差分是使得序列平稳化的途径之一。完成差分后，ARIMA 模型的建立与 ARMA 类似，确保了序列的稳定性。在建模过程中，通过分析样本自相关(ACF)和偏自相关(PACF)确定模型参数。之后，利用最大似然等方法估计参数，并通过残差分析验证模型的适应性。

2.3. 随机森林模型

随机森林算法构建于决策树之上，通过集成多棵决策树的预测结果，对于回归任务，它采用所有树的结果平均值作为最终预测；对于分类任务，则采用多数投票原则确定最终类别。随机森林中的每棵决策树都是独立的，相互之间没有直接联系。当进行分类时，新样本会经过森林中每棵树的独立判断和分类，随机森林将选择出现次数最多的分类结果作为最终输出。以下是随机森林算法的实施步骤：

1) 在包含 N 个样本的数据集中，进行有放回的随机抽样，每次抽取一个样本后放回，共进行 N 次，以这批样本训练一棵决策树，这些样本将构成树的根节点。

2) 假设每个样本具有 M 个属性，在决策树节点分裂过程中，随机从这 M 个属性中抽取 m 个(m 远小于 M)，并基于某种标准(例如信息增益最大化)从这 m 个属性中挑选出最佳属性用于节点分裂。

3) 在决策树的生长过程中，不断重复步骤 2，直到节点无法进一步分裂，或达到预设的停止条件，如树的最大深度或叶子节点数量的限制。

重复步骤(1)至(3)，生成大量的决策树，这些树共同构成了随机森林。

本研究采用农产品数据，通过相关性分析确定了影响平均价格的关键因素，并将其作为解释变量，以平均价格作为预测变量，建立了随机森林预测模型。模型的构建使用了 Sklearn. Ensemble 库中的 Random Forest Regressor 方法。

3. 实证分析

数据收集自指定接口，通过 Python 脚本自动获取并存储至 CSV 文件。数据预处理包括清洗无效数据、填充缺失值等。指标体系构建基于理论支撑，选择了关键指标进行全面分析。数据可视化通过 Matplotlib 库实现，包括趋势图、直方图和箱线图。模型构建与结果分析详细展示了线性回归、ARIMA 模型拟合和随机森林回归模型的训练过程和预测效果。本部分将详细描述数据的收集、处理、指标体系构建、数据可视化以及模型的构建和结果分析过程。

3.1. 数据收集与预处理

数据收集是实证分析的第一步。通过编写 Python 脚本，利用 Requests 库从指定的接口自动获取农产品价格数据。数据包括 2022 年 1 月 1 日到 2024 年 6 月 24 日的商品名称、分类、平均价、最低价、最高价、发布时间、日期和产地等字段。数据收集后，使用 Pandas 库进行数据清洗和预处理，包括去除无效数据、填充缺失值、转换数据类型以及将时间戳转换为日期格式等。

3.2. 指标体系构建

在数据预处理的基础上，构建了一个包含多个关键指标的体系，以全面分析农产品价格的波动特征。这些指标包括但不限于平均价、最低价、最高价等。每个指标的选择都有其理论依据，例如平均价可以反映市场价格的整体水平，而最高价和最低价则可以揭示价格的波动范围。处理完后的数据出存在 csv 文件中打开的图如图 1 所示。数据部分展示如表 1。

名称	分类	平均价	最低价	最高价	时期	日期
大白菜	蔬菜	0.48	0.4	0.55	2024/6/23 0:00	2024/6/23
大白菜	蔬菜	0.48	0.4	0.55	2024/6/22 0:00	2024/6/22
大白菜	蔬菜	0.38	0.3	0.45	2024/6/21 0:00	2024/6/21
大白菜	蔬菜	0.38	0.3	0.45	2024/6/20 0:00	2024/6/20
大白菜	蔬菜	0.48	0.45	0.5	2024/6/19 0:00	2024/6/19
大白菜	蔬菜	0.48	0.45	0.5	2024/6/18 0:00	2024/6/18
大白菜	蔬菜	0.4	0.3	0.5	2024/6/17 0:00	2024/6/17
大白菜	蔬菜	0.35	0.3	0.4	2024/6/16 0:00	2024/6/16
大白菜	蔬菜	0.35	0.3	0.4	2024/6/15 0:00	2024/6/15
大白菜	蔬菜	0.35	0.3	0.4	2024/6/14 0:00	2024/6/14
大白菜	蔬菜	0.35	0.3	0.4	2024/6/13 0:00	2024/6/13
大白菜	蔬菜	0.35	0.3	0.4	2024/6/12 0:00	2024/6/12
大白菜	蔬菜	0.35	0.3	0.4	2024/6/11 0:00	2024/6/11
大白菜	蔬菜	0.35	0.3	0.4	2024/6/10 0:00	2024/6/10
大白菜	蔬菜	0.35	0.3	0.4	2024/6/9 0:00	2024/6/9
大白菜	蔬菜	0.4	0.3	0.5	2024/6/8 0:00	2024/6/8
大白菜	蔬菜	0.4	0.3	0.5	2024/6/7 0:00	2024/6/7
大白菜	蔬菜	0.4	0.3	0.5	2024/6/6 0:00	2024/6/6
大白菜	蔬菜	0.45	0.4	0.5	2024/6/5 0:00	2024/6/5
大白菜	蔬菜	0.45	0.4	0.5	2024/6/4 0:00	2024/6/4
大白菜	蔬菜	0.45	0.4	0.5	2024/6/3 0:00	2024/6/3
大白菜	蔬菜	0.35	0.25	0.45	2024/6/2 0:00	2024/6/2
大白菜	蔬菜	0.33	0.25	0.4	2024/6/1 0:00	2024/6/1
大白菜	蔬菜	0.33	0.25	0.4	2024/5/31 0:00	2024/5/31
大白菜	蔬菜	0.4	0.35	0.45	2024/5/30 0:00	2024/5/30
大白菜	蔬菜	0.4	0.35	0.45	2024/5/29 0:00	2024/5/29
大白菜	蔬菜	0.4	0.35	0.45	2024/5/28 0:00	2024/5/28

大白菜	蔬菜	0.4	0.35	0.45	2024/5/27 0:00	2024/5/27
大白菜	蔬菜	0.45	0.4	0.5	2024/5/26 0:00	2024/5/26
大白菜	蔬菜	0.35	0.3	0.4	2024/5/25 0:00	2024/5/25
大白菜	蔬菜	0.35	0.3	0.4	2024/5/24 0:00	2024/5/24
大白菜	蔬菜	0.38	0.3	0.45	2024/5/23 0:00	2024/5/23
大白菜	蔬菜	0.4	0.35	0.45	2024/5/22 0:00	2024/5/22
大白菜	蔬菜	0.43	0.35	0.5	2024/5/21 0:00	2024/5/21
大白菜	蔬菜	0.43	0.35	0.5	2024/5/20 0:00	2024/5/20
大白菜	蔬菜	0.43	0.35	0.5	2024/5/19 0:00	2024/5/19
大白菜	蔬菜	0.48	0.35	0.6	2024/5/18 0:00	2024/5/18
大白菜	蔬菜	0.5	0.4	0.6	2024/5/17 0:00	2024/5/17
大白菜	蔬菜	0.5	0.4	0.6	2024/5/16 0:00	2024/5/16
大白菜	蔬菜	0.55	0.45	0.65	2024/5/15 0:00	2024/5/15
大白菜	蔬菜	0.55	0.45	0.65	2024/5/14 0:00	2024/5/14
大白菜	蔬菜	0.55	0.45	0.65	2024/5/13 0:00	2024/5/13

Figure 1. Data presentation chart
图 1. 数据展示图

Table 1. Partial data presentation
表 1. 部分数据展示

名称	分类	平均价	最低价	最高价	时期	日期	时间	产地
大白菜	蔬菜	0.35	0.3	0.4	2024/6/9 0:00	2024/6/9	0:00:00	冀
大白菜	蔬菜	0.4	0.3	0.5	2024/6/8 0:00	2024/6/8	0:00:00	冀
大白菜	蔬菜	0.4	0.3	0.5	2024/6/7 0:00	2024/6/7	0:00:00	冀
大白菜	蔬菜	0.4	0.3	0.5	2024/6/6 0:00	2024/6/6	0:00:00	冀
大白菜	蔬菜	0.45	0.4	0.5	2024/6/5 0:00	2024/6/5	0:00:00	冀
大白菜	蔬菜	0.45	0.4	0.5	2024/6/4 0:00	2024/6/4	0:00:00	冀

3.3. 数据可视化

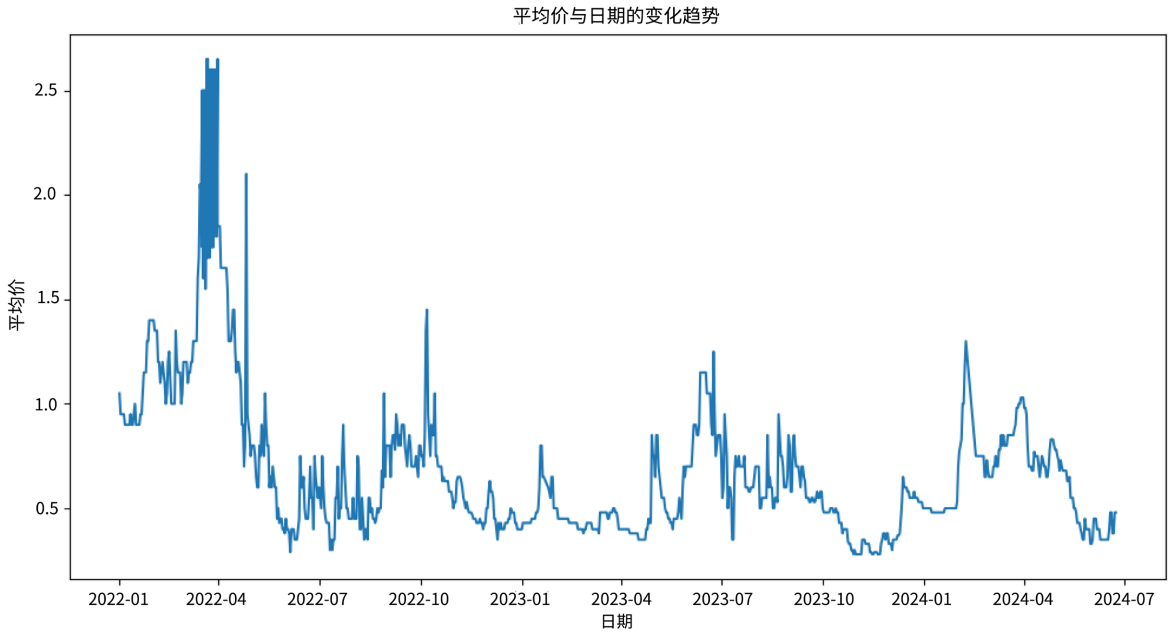

Figure 2. Trends in average price versus date
图 2. 平均价与日期的变化趋势

为了更直观地理解数据，使用 Matplotlib 库对数据进行可视化分析。绘制了平均价随日期变化的趋势图，以及最高价和最低价的变化趋势图。此外，还通过直方图和箱线图展示了价格的分布情况，这些图表有助于初步理解价格数据的波动性和分布特征。平均价与日期关系如图 2，最高价与最低价之间如图 3，以及平均价的分布情况如图 4，图 5 所示

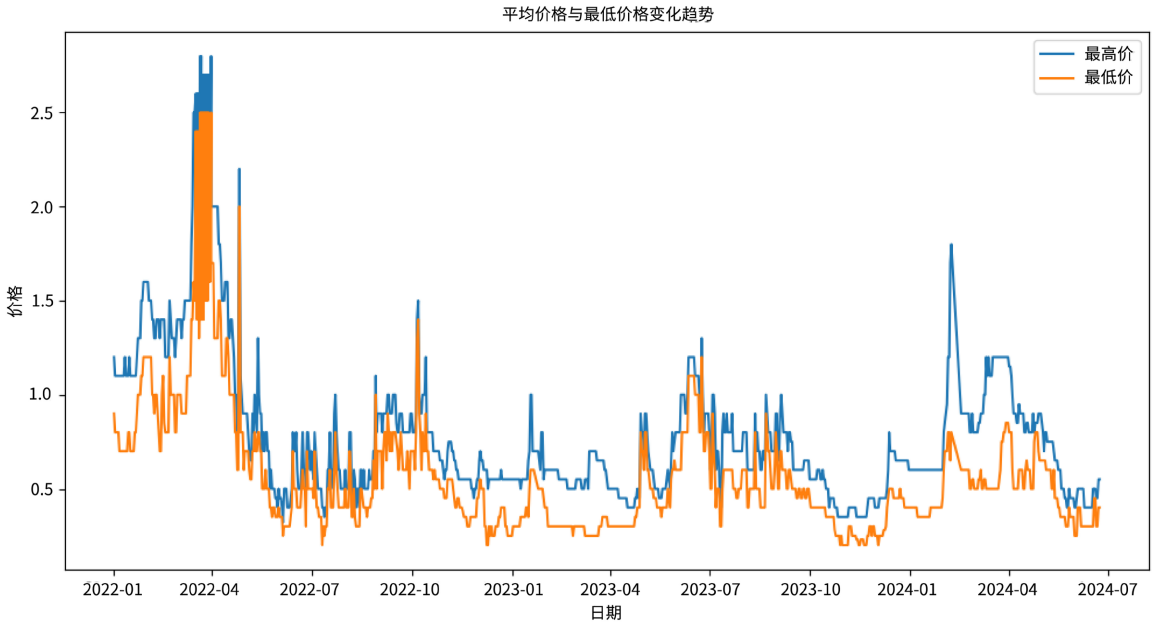

Figure 3. Trends in high and low prices
图 3. 最高价与最低价变化趋势

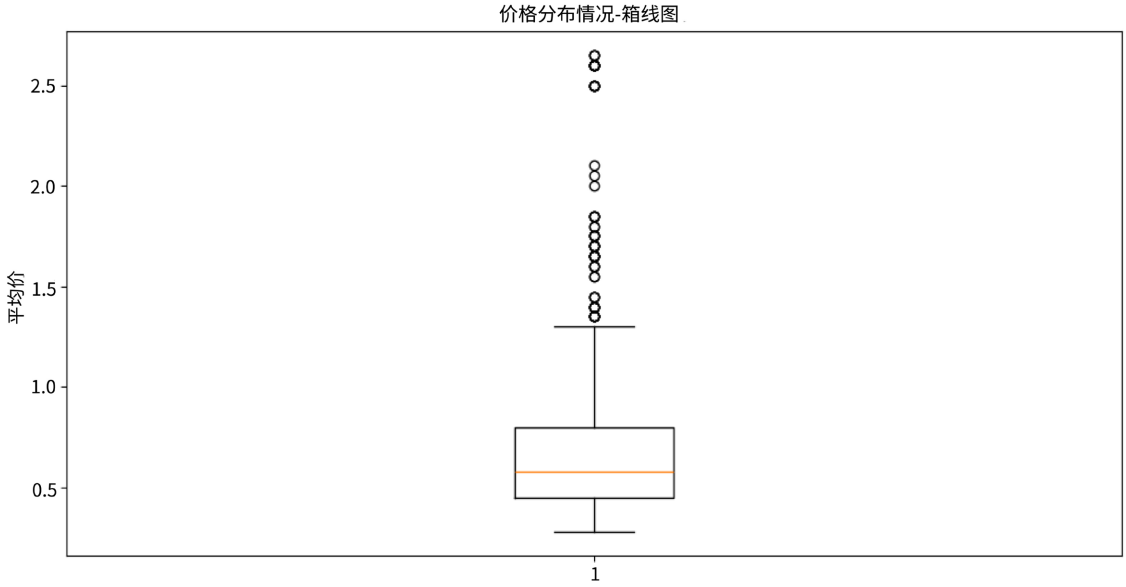

Figure 4. Box plot
图 4. 箱线图

根据图 5 可知，大白菜的价格多分布在 0.25 到 1 之间，全年来说是以先上升后下降的趋势，采取不同模型对其价格变化趋势进行研究。

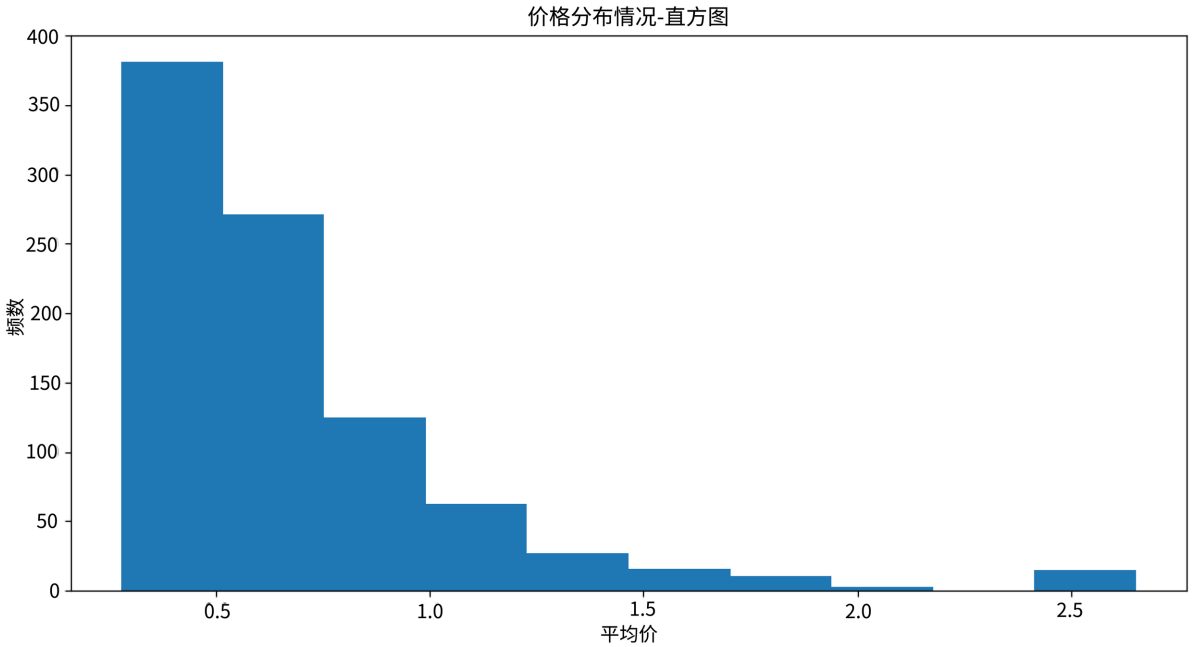

Figure 5. Histogram
图 5. 直方图

3.4. 模型构建与结果分析

在数据可视化的基础上，进一步构建了多个统计模型来分析和预测农产品价格。

3.4.1. 线性回归模型

选择日期与产地为自变量，价格为因变量。对产地进行独热编码，确保模型的准确性和可靠性。通过输出 ols. Summary，得到结果如图 6 所示。

	coef	std err	t	P>|t|	[0.025	0.975]
const	7.1352	0.914	7.804	0.000	5.341	8.930
日期	-4.026e-09	5.86e-10	-6.864	0.000	-5.18e-09	-2.87e-09
产地_冀	0.3019	0.076	3.995	0.00	0.154	0.450
产地_冀晋	0.1470	0.114	1.293	0.196	-0.076	0.370
产地_冀蒙	0.2277	0.101	2.254	0.024	0.029	0.426
产地_冀鄂	0.0363	0.091	0.401	0.689	-0.142	0.214
产地_冀鄂鲁	0.0780	0.096	0.814	0.416	-0.110	0.266
产地_冀鲁	0.1886	0.079	2.380	0.018	0.033	0.344
产地_冀鲁辽鄂	0.6387	0.064	9.991	0.000	0.513	0.764
产地_冀鲁鄂	0.3727	0.090	4.136	0.000	0.196	0.550
产地_鄂冀鲁	0.0211	0.100	0.212	0.833	-0.175	0.217
产地_鲁云	1.1531	0.174	6.608	0.000	0.811	1.496
产地_鲁云浙	0.9097	0.098	9.283	0.000	0.717	1.102
产地_鲁冀津	0.5590	0.140	3.985	0.000	0.284	0.834
产地_鲁鄂桂	1.1784	0.107	11.052	0.000	0.969	1.388
产地_鲁鄂桂冀	1.3228	0.122	10.835	0.000	1.083	1.562

Figure 6. Summary chart
图 6. Summary 图

观察图 6 可以发现不同产地作为不同的自变量出现，部分 p 值大于 0.05。为确保模型的显著性，对 p 值大于 0.05 的自变量进行逐步剔除，得到结果如图 7 所示。

剔除后的模型表达式：

OLS Regression Results

Dep. Variable:	平均价	R-squared:	0.412
Model:	OLS	Adj. R-squared:	0.404
Method:	Least Squares	F-statistic:	57.09
Date:	Mon,15 Jul 2024	Prob(F-statistic):	1.26e-95
Time:	20:53:07	Log-Likelihood:	-208.63
No. Observations:	910	AIC:	441.3
Df Residuals:	898	BIC:	499.0
Df Model:	11		
Covariance Type:	nonrobust		

Figure 7. Improved Summary chart
图 7. 改良后的 Summary 图

根据图 7 所示 $R^2 = 0.412$ 。表明该模型准确精度弱，此模型被排除。

3.4.2. ARIMA 模型

先通过 ADF 检验，得到结果如图 8 所示。

Augmented Dickey-Fuller Test: Original Series

ADF test statistic	-3.853499
p-value	0.002405
# lags used	3.000000
# observations	724.000000
critical value (1%)	-3.439414
critical value (5%)	-2.865540
critical value (10%)	-2.568900
Strong evidence against the null hypothesis	
Reject the null hypothesis	
Data has no unit root and is stationary	

Figure 8. ADF test
图 8. ADF 检验

由图 8 可得，ADF 检验结果显示，原始时间序列的 ADF 统计量为−3.853499，p 值为 0.002405。由于 p 值小于 0.05，拒绝原假设，表明原始时间序列是平稳的，因此差分为 0 阶。

根据 ACF 图和 PACF 图对模型进行定阶具有一定的主观性，因此本文根据 AIC 准则和 BIC 准则，来确定模型的阶数。本文确定 p，q 的范围[0, 3]，通过循环网格搜索所有组合的 AIC 和 BIC 的值，得到结果见表 2。

Table 2. AIC and BIC values of the portfolio
表 2. 组合的 AIC 与 BIC 值

模型	AIC	BIC
ARIMA (0, 0, 0)	−343.2	−334.0
ARIMA (0, 0, 1)	−934.9	−921.2

续表

ARIMA (0, 0, 2)	−1230.6	−1212.2
ARIMA (0, 0, 3)	−1324.1	−1301.2
ARIMA (1, 0, 0)	−1589.0	−1575.2
ARIMA (1, 0, 1)	−1594.4	−1576.0
ARIMA (1, 0, 2)	−1613.9	−1591.0
ARIMA (1, 0, 3)	−1617.0	−1589.5
ARIMA (2, 0, 0)	−1591.6	−1573.3
ARIMA (2, 0, 1)	−1610.0	−1587.0
ARIMA (2, 0, 2)	−1614.1	−1586.6
ARIMA (2, 0, 3)	−1617.3	−1585.2
ARIMA (3, 0, 0)	−1607.1	−1584.2
ARIMA (3, 0, 1)	−1616.8	−1589.2
ARIMA (3, 0, 2)	−1618.5	−1586.4
ARIMA (3, 0, 3)	−1619.4	−1582.7

通过对比各模型参数的 AIC 和 BIC 的值可以发现，0 阶差分的 ARMA 模型为 ARMA (1, 2)时，模型认为模型参数的 AIC 和 BIC 的值相对最小，所以当时为最佳 $p=1$, $q=1$阶数。

根据选择的 ARIMA (1, 0, 2)模型，得到 ARIMA 模型结果如图 9。

考虑产地的 ARIMA 模型结果:

SARIMAX Results

Dep. Variable:	平均价	No. Observations:	728
Model:	ARIMA(1,0,2)	Log Likelihood	811.969
Date:	Mon,15 Jul 2024	AIC	-1613.939
Time:	23:11:49	BIC	-1590.987
Sample:	0	HQIC	-1605.083
	-728		
Covariance Type:	opg		

Figure 9. ARIMA model
图 9. ARIMA 模型

3.4.3. 随机森林模型

Table 3. Table of cross-scores for parameters
表 3. 参数的交叉得分表

参数	交叉得分
N_Estimators: 100, Max_Depth: None, cv: 3	0.9255
……	……
N_Estimators: 300, Max_Depth: 20, cv: 7	0.9286
……	……
N_Estimators: 400, Max_Depth: 20, cv: 7	0.9277

随机森林模型划分训练集和测试集，测试集占原有数据的 20%。为保证模型结果的可重复性，将随机种子设置为 0。设定 Parma_Grid 的 N_Estimators [100, 200, 300, 400]和 Max_Depth [None, 5, 10, 15, 20]以及使用 Grid Search CV Cv_Range [3, 5, 7]进行网格搜索，交叉验证找到最优参数如表 3，得到最优模型参数为[N_Estimators: 300, Max_Depth: 20, cv: 7]，利用最优的参数进行模型预测并评估了模型的预测准确性如图 10。

模型类型：随机森林回归
参数：
n_estimators: 300
max_depth:20
性能指标：
均方误差（MSE）：0.0064
均方根误差（RMSE）：0.0799
R^2 得分：0.9575

Figure 10. Random forest model
图 10. 随机森林模型

通过图 10，可以看出 R^2 为 0.96，均方差为 0.006，该模型拟合效果较好。

3.4.4. 三种模型准确对比分析

分别对比了线性回归模型，随机森林模型，ARIMA 模型的准确度，结果如表 4 所示。通过对比发现随机森林模型效果优于 ARIMA 模型，ARIMA 模型优于线性回归模型。

Table 4. Comparative analysis table of the accuracy of each model
表 4. 各模型准确度对比分析表

模型	准确度
线性回归模型	0.412
ARIMA 模型	0.512
随机森林模型	0.957

3.4.5. FLASK 框架实现

Figure 11. Operation flow chart
图 11. 操作流程图

为了将本研究转化为可直观查看与交互的网页，降低读者学习以及使用成本，本文利用 FLASK 框架构造了一个轻量级的 WEB。利用 HTML 语言编写前端的 Index. Html，用于展示页面。然后用 Result. Html 进行后端处理。采用了 Css，Javascript，Html 技术，设计网页的按钮规格，颜色，展示动态等使页面更加美观简洁。用户只需选择日期以及预测的产地后点击预测后会接入前文调试好的随机森林模型，进行预测而后在页面返回给用户预测值。界面简单且操作简洁。具体操作如图 11 所示。

4. 结论

本研究成功构建并对比了线性回归，随机森林，ARIMA 模型在农产品价格预测中的有效性。不同的模型在捕捉价格变化趋势不同的效果，其中随机森林模型最好。为市场参与者提供了可靠的决策支持。研究的局限性在于数据集的时间范围和多样性，未来研究将扩大数据规模，探索更多机器学习算法，并结合经济学理论进行深入分析。本研究通过多元统计分析方法，结合时间序列和机器学习技术，对农产品价格趋势进行了深入的实证分析和预测。

本文利用 Request 库从新发地发起请求得到对象后进行 Xpath 获取到 2022 年 1 月至 2024 年 6 月每一种蔬菜，肉类，调料，水产的价格数据源，进行筛选获得 2022 年 1 月至 2024 年 6 月每天大白菜的发售价，后续进行数据重复的进行删除，这些操作为后续分析打下了坚实的基础。研究结果表明线性回归模型拟合效果不好准确率只有 41.2%。ARIMA 模型在分析平均价格趋势方面表现出了良好的性能，准确率达到为 51.2%，精度高于线性回归模型。随机森林回归模型，在参数调优后，预测准确性高达 95.7%，高于其他两个模型。

尽管本研究在农产品价格预测方面取得了一定的成果，但仍存在一些局限性：模型的选择和参数调优可能需要更多的实验来验证其稳健性；实证分析主要集中在统计和机器学习方法，对于经济学理论和市场机制的深入探讨不足。针对现有研究的局限性，提出以下未来研究方向：扩大数据集的规模和多样性，以提高模型的泛化能力和预测的准确性；探索更多的机器学习算法，如深度学习模型，以进一步提高预测性能；结合经济学理论，深入分析影响农产品价格的宏观经济因素和市场机制；考虑季节性因素和政策变动对农产品价格的影响，构建更为复杂的预测模型。

综上所述，本研究通过多元统计分析和机器学习方法，为农产品价格预测提供了一种新的视角和方法。研究成果不仅对市场参与者具有实际指导意义，也为相关领域的学术研究提供了参考。未来，将继续优化模型，拓展研究范围，并深化对市场机制的理解。

基金项目

重庆对外经贸学院科学研究项目(KYKJ202205)。

参考文献

[1] 刘峰, 王儒敬, 李传席. ARIMA 模型在农产品价格预测中的应用[J]. 计算机工程与应用, 2009, 45(25): 238-239+248.

[2] 张瑞瑜. 基于舆情分析的小宗农产品价格预测及系统原型设计[D]: [硕士学位论文]. 哈尔滨: 东北农业大学, 2022.

[3] 郭锋. 基于大数据的大蒜价格预测及可视化研究[D]: [硕士学位论文]. 泰安: 山东农业大学, 2019.

[4] 马小菁. 基于时间序列分析的山东大葱价格预测研究[D]: [硕士学位论文]. 烟台: 烟台大学, 2021.

[5] 贺本岚. 股票价格预测的最优选择模型[J]. 统计与决策, 2008(6): 135-137.

[6] 王燕. 时间序列分析: 基于 R [M]. 北京: 中国人民大学出版社, 2020: 6.