

分类号 C8/415  
UDC

密级 公开  
编号 10741



# 硕士学位论文

(专业学位)

论文题目 基于机器学习的我国粮食产量预测研究

研究生姓名: 肖文

指导教师姓名、职称: 庞智强、教授

学科、专业名称: 应用统计

研究方向: 经济统计应用

提交日期: 2024年6月3日

## 独创性声明

本人声明所呈交的论文是我个人在导师指导下进行的研究工作及取得的研究成果。尽我所知，除了文中特别加以标注和致谢的地方外，论文中不包含其他人已经发表或撰写过的研究成果。与我一同工作的同志对本研究所做的任何贡献均已在论文中作了明确的说明并表示了谢意。

学位论文作者签名： 张 签字日期： 2024.6.3  
导师签名： 张 签字日期： 2024.6.3  
校外导师签名： \_\_\_\_\_ 签字日期： \_\_\_\_\_

## 关于论文使用授权的说明

本人完全了解学校关于保留、使用学位论文的各项规定， 同意（选择“同意” / “不同意”）以下事项：

1. 学校有权保留本论文的复印件和磁盘，允许论文被查阅和借阅，可以采用影印、缩印或扫描等复制手段保存、汇编学位论文；

2. 学校有权将本人的学位论文提交至清华大学“中国学术期刊（光盘版）电子杂志社”用于出版和编入 CNKI《中国知识资源总库》或其他同类数据库，传播本学位论文的全部或部分內容。

学位论文作者签名： 张 签字日期： 2024.6.3  
导师签名： 张 签字日期： 2024.6.3  
校外导师签名： \_\_\_\_\_ 签字日期： \_\_\_\_\_

# **Research on Grain Production Prediction Based on Machine Learning**

**Candidate : Xiao Wen**

**Supervisor: Pang Zhiqiang**

## 摘要

在面对全球粮食市场的不确定性和挑战，粮食产量预测变得至关重要。粮食产量预测不仅可以帮助我们及时应对自然灾害、气候变化等因素对农业生产的影响，还能为国家粮食安全战略的制定提供关键信息。通过粮食产量预测，我们能够更好地规划农业生产、合理配置资源，进而保障粮食供应，维护国家经济的稳定和人民的福祉。因此，加强粮食产量预测工作，提高预测准确性和可靠性，对于确保粮食安全、促进农业可持续发展具有重要意义。

本文通过收集 1949 年至 2022 年的我国粮食产量数据和影响因素等数据，结合机器学习算法的强大分析能力，建立了粮食产量预测模型。接着为了比较不同的数据集对模型进行预测的效果，继续收集了 1949 年至 2022 年江西省粮食产量数据和影响因素等数据用来训练和评估模型在面对不同数据特征和分布时的稳健性和可靠性。结果显示，两种不同的数据集在模型训练与测试中均表现良好。在这两个数据集上，BP 神经网络和随机森林模型表现出较好的训练和测试效果，其次是支持向量回归模型和长短期记忆神经网络模型。

本文提出采用 Stacking 集成算法，将 RF、BP、SVR 以及 LSTM 模型进行融合，构建 Stacking 集成模型。将构建后的 Stacking 算法集成模型与以往的平均法组合模型进行预测效果对比分析，并采用交叉验证的方法来进一步评估模型的稳定性。实验结果表明，相较于传统的平均法集成模型与单一机器学习模型，Stacking 算法构建的集成模型均展现出优异的预测效果。在 5 折交叉验证中，RF-BPNN-SVR-Stacking 模型预测效果显著强于其它模型；在 10 折交叉验证中，RF-BPNN-SVR-LSTM-Stacking 模型的预测实验误差最小，预测效果最好。RF-BPNN-SVR-Stacking、RF-BPNN-SVR-LSTM-Stacking 模型相较于传统的平均法集成模型在不同的交叉验证方式下均展现出优异的预测效果，为粮食产量预测提供了新的优化思路，有望提升预测精度。

**关键词:**机器学习 粮食产量预测 Stacking 集成模型 交叉验证

# Abstract

In the face of uncertainty and challenges in the global grain market, predicting grain production has become crucial. Grain production forecasting not only helps us timely respond to factors such as natural disasters and climate change affecting agricultural production, but also provides key information for formulating national food security strategies. Through grain production forecasting, we can better plan agricultural production, allocate resources reasonably, thereby ensuring food supply, maintaining economic stability, and safeguarding the well-being of the people. Therefore, strengthening grain production forecasting work and improving forecasting accuracy and reliability are of great significance for ensuring food security and promoting sustainable agricultural development.

This paper establishes grain production prediction models by collecting data on grain production from 1949 to 2022 in China, along with other influencing factors, and leveraging the powerful analytical capabilities of machine learning algorithms. To compare the effects of different datasets on model predictions, grain production data and influencing factors from Jiangxi Province from 1949 to 2022 were collected to train and evaluate the robustness and reliability of the models when facing different data features and distributions. The results show that both datasets performed well in model training and testing. On both datasets, the BP neural network and random forest models exhibited good training and testing performance, followed by the SVR model and long short-term memory neural network model.

This paper proposes the use of Stacking ensemble learning algorithm to integrate RF, BP, SVR, and LSTM models, constructing a Stacking ensemble model. The predictive performance of the constructed Stacking ensemble model is compared with that of traditional averaging ensemble models, and the stability of the models is further evaluated using cross-validation. Experimental results demonstrate that, compared to traditional averaging ensemble models and single machine learning models, the Stacking ensemble models exhibit excellent predictive performance. In 5-fold cross-

validation, the RF-BPNN-SVR-Stacking model significantly outperformed other models; in 10-fold cross-validation, the RF-BPNN-SVR-LSTM-Stacking model showed the smallest experimental prediction error and the best predictive performance. The RF-BPNN-SVR-Stacking and RF-BPNN-SVR-LSTM-Stacking models demonstrate excellent predictive performance under different cross-validation methods compared to traditional averaging ensemble models, providing new optimization ideas for grain production prediction and potentially improving prediction accuracy.

**Keywords:** Machine Learning; Grain Production Prediction; Stacking Ensemble Model; Cross-Validation.

# 目 录

<b>1 绪论 .....</b>	<b>1</b>
1.1 研究背景与意义 .....	1
1.2 国内外研究现状 .....	2
1.2.1 国外研究现状 .....	2
1.2.2 国内研究现状 .....	4
1.2.3 文献述评 .....	6
1.4 研究框架 .....	7
1.5 论文创新点 .....	9
<b>2 相关概念和理论基础 .....</b>	<b>10</b>
2.1 粮食产量预测的相关理论 .....	10
2.1.1 粮食产量的相关概念 .....	10
2.1.2 粮食产量的影响因素 .....	11
2.2 机器学习相关算法基本原理 .....	13
2.2.1 BP 神经网络 .....	13
2.2.2 支持向量回归 .....	14
2.2.4 随机森林 .....	15
2.2.1 长短期记忆神经网络 LSTM .....	16
2.3 Stacking 算法与模型融合 .....	18
2.4 交叉验证 .....	19
2.5 本章小结 .....	20
<b>3 基于单一模型的粮食产量预测 .....</b>	<b>21</b>
3.1 基于 LSTM 模型的粮食产量预测分析 .....	21
3.1.1 数据来源及评价指标 .....	21
3.1.2 LSTM 模型建模过程 .....	22

3.2.3 模型预测与结果分析 .....	23
3.2 基于 BP 神经网络模型的粮食产量预测分析 .....	23
3.2.1 BP 神经网络模型构建 .....	24
3.2.2 BP 神经网络模型训练与评估 .....	24
3.3 支持向量回归模型的粮食产量预测 .....	25
3.4 基于随机森林模型的粮食产量预测分析 .....	27
3.5 本章小结 .....	29
<b>4 基于 Stacking 算法的粮食产量预测 .....</b>	<b>30</b>
4.1 Stacking 集成模型构建 .....	31
4.2 Stacking 集成模型训练与评估 .....	32
4.3 使用平均法的集成模型 .....	35
4.4 本章小结 .....	37
<b>5 粮食产量预测模型的对比分析 .....</b>	<b>38</b>
5.1 预测模型的对比与分析 .....	38
5.2 对比数据集预测结果 .....	40
5.2 使用交叉验证的预测模型对比与分析 .....	41
5.3 本章小结 .....	43
<b>6 总结与展望 .....</b>	<b>44</b>
6.1 本文总结 .....	44
6.2 本文展望 .....	44
<b>参考文献 .....</b>	<b>46</b>
<b>致谢 .....</b>	<b>50</b>

# 1 绪论

## 1.1 研究背景与意义

粮食安全是国家发展的重中之重，历来受到广泛关注。而随着我国经济的快速发展和人口的持续增长，粮食安全问题也越来越凸显。在粮食安全中粮食的充足供应一直是我国社会稳定和经济发展的基石。作为人类生存和发展的基本需求之一，充足的粮食产量直接关系到国家和人民的生计和安全。稳定的粮食产量不仅能够满足人们的日常生活需求，也是维护社会稳定、促进经济繁荣的重要基础。在全球化和气候变化的背景下，粮食产量的稳定性和增长性显得尤为重要。通过确保充足的粮食供应，可以缓解饥饿问题，提高人民的生活水平，促进社会和谐稳定。因此，保障粮食产量的稳定增长，不仅是国家发展战略的重要组成部分，也是实现可持续发展和构建和谐社会的必然选择。在我国的发展历程中，粮食产量问题一直是我国经济发展中的一个重要议题，因为粮食的供应充足与否直接关系到国家经济的稳定和社会的进步。党和政府长期以来也高度重视粮食安全，将保障粮食供应视为国家安全的核心任务之一。自 1996 年国务院发布《中国的粮食安全问题》白皮书，提出了“立足国内资源、实现粮食基本自给”的粮食战略，到 2023 年中央再次将粮食稳产保供作为治国理政的首要任务，着重确保全国粮食产量稳定在 1.3 亿斤以上。这些措施充分彰显了党和国家对粮食问题的高度重视，将解决吃饭问题置于治理国家的首要位置。

随着科技的发展和应用，我国农业生产正逐步向现代化转型。农业机械化、智能化、信息化等技术的广泛应用，提高了农业生产效率和质量。但与此同时，我们也面临着新的挑战，全球市场的变化和波动会影响粮食的价格与供需关系，再加上气候的变化会对农业造成不确定性的挑战。因此，准确预测粮食产量具有重要意义，准确的粮食产量预测有助于应对国际市场的变化，降低对进口粮食的依赖，同时也能帮助农民和政府及时应对气候变化的影响，采取适当的调整和应对措施来保障国内市场供应。为了保障我国粮食安全和稳定供应，我们需要提前对粮食生产形势进行判断。粮食预测在这一领域发挥着重要作用，它可以相对精

准地预估未来短期粮食产量，为决策者提供重要参考依据，例如制定粮食政策、调整农业生产计划、安排粮食进出口等。通过粮食预测，决策者可以及时了解粮食供需状况，采取相应措施，保障国家粮食安全和市场稳定。

在保障粮食安全和维护粮食产量供应稳定方面，机器学习技术起着重要作用。通过将机器学习技术应用于粮食生产过程中，我们能更加科学的对粮食产量进行规划。例如，机器学习可以分析粮食储存环境中的温湿度、气氛气体成分等数据，预测粮食储存中可能出现的问题，如霉变、虫害等，及时采取措施进行预防和处理。也可以通过分析历史数据和实时信息，优化粮食运输路线和运输计划，提高运输效率，降低成本，并且可以预测潜在的交通拥堵和运输风险，采取相应的应对措施。并且在农业灾害预测和防范中机器学习技术可以利用大数据和传感器信息对农业灾害进行预测和监测，包括洪涝灾害、干旱、病虫害等。通过分析大量的数据，可以及时发现灾害风险，并采取预防措施，减少灾害损失。因此，在农业领域中发展机器学习技术用于粮食产量预测具有重要现实意义。当前，粮食生产面临土地资源压力、水资源紧缺、气候变化影响以及人口增长压力等诸多挑战。因此，提高粮食生产的稳定性以及高效性是当前亟待解决的重要问题。而机器学习技术因能够处理大规模的农业数据，并从中挖掘出隐藏的模式和关联。并且可以根据不同地区和作物的特性，为农民提供个性化的种植方案，成为确保粮食安全的重要手段。

## 1.2 国内外研究现状

### 1.2.1 国外研究现状

在农产量的预测，国外研究人员通常会将气象数据以及遥感数据引入到粮食产量预测中。这些数据可以提供更丰富的信息，例如土壤水分、气温、降水量等，有助于提高预测模型的准确性和可靠性。而获得这些数据使用的主要工具气象预测，以及卫星监测。因此他们在对粮食产量进行预测时主要使用的方法是气象预测法，卫星遥感技术预测以及使用生长模拟的方法来预测粮食产量。气象预测法是一种利用气象数据和气象模型进行预测的方法，用于预测未来一段时间内的气

象变化和天气情况。在粮食产量预测中，气象预测法通常通过分析气象因素对作物生长和产量的影响，从而对未来的粮食产量进行预测。然而当前世界科学技术水平下无法对长时间段的气象进行准确预测，因此在对粮食产量进行预测时无法获得与之相关的降水量、温度等数据。因此 Ramesh (2016) 指出气象预测技术也存在着一定的局限性，即无法准确进行中长期预测。为了解决获取数据问题，研究人员开始运用遥感技术来获取粮食产量相关的数据，但是遥感技术还处在发展初期阶段，还未达到成熟的地步，因此遥感技术在粮食产量预测中的应用还未全部开放。在生物学领域中，研究人员会使用模拟的方法来探究对生物生长的影响因素。但是生长模拟法只局限于小域范围，在大范围预测模拟上使用成本太高。

尽管面对着诸多困难，国外研究人员依然热衷于对粮食产量预测进行研究，并且取得了一些研究成果。比如国外学者将农作物生长检测系统进行改进，然后基于改进后的 CGMS、原 CGMS 分别对欧洲国家主要农作物的生长状况进行检测，并采用农业气候指标对模型进行冲击。实验结果表明改进后的 CGMS 能够更好的对农作物产量进行预测 (Pagani, 2017)。尽管 CGMS 在作物监测和管理方面具有许多优势，但也存在一些缺陷，比如，CGMS 需要大量的遥感数据和地面观测数据作为输入，而且遥感数据通常具有高分辨率和大数量，因此需要进行复杂的数据处理和分析，包括图像处理、特征提取、分类和建模等过程。于是有学者在获取高分辨率数据上进行创新，比如 Meshesha (2018) 利用 Sentinel-2 遥感影像数据，然后开发了一个新的预测系统对农作物产量进行预测，结果发现使用高分辨数据对粮食产量预测能够预测的更精准。高分辨率数据虽然对粮食产量预测具有关键作用，但是其大量获取依然有一定的难度，为了获取大量高分辨率数据，Awad (2019) 基于能量守恒定律，构建了一种新模型，并获取当地主要农作物大量数据来验证这个新模型的有限性，实验结果发现，使用大量高分辨率数据对粮食产量预测的预测误差显著性减小。农业干旱对于粮食的产出具有重要影响，于是有学者开始构建农业干旱指数，而在构建农业干旱指数时有基于气象学方法，也有基于卫星遥感技术。为了验证这两种方法构建的农业干旱指数在粮食产量预测中的有效性，Garcia-León (2019) 等人的团队使用西班牙地区粮食产量

数据，然后构建气候统计模型，用该模型来研究这两种指数对粮食产量的影响。实验结果表明这两个指数能够有效的解释粮食产出。

随着芯片的不断升级迭代，计算机在处理越来越多的数据也越来越强大，机器学习技术开始新起，国外一些学者开始转向于使用机器学习与卫星监测，气象动力学相结合来预测粮食产量。例如 Liu (2017) 针对当地小麦生长，使用卫星监测来获取小麦生长影像，并通过机器学习技术对获取的小麦生长影像数据进行识别与分类，然后再利用其特征进行小麦产量的预估。通过该项研究发现使用机器视觉技术能够很好的估计小麦生长的密度，而植物生长密度可以很好反映出该地区的小麦产量。国外学者为了验证机器学习对于粮食产量预测适用性，将 LSTM 模型、卷积神经网络模型与遥感技术以及气象预测技术相结合构建成两个不同的机器学习模型，接着使用基于 LSTM 的组合模型来处理气象与遥感数据，用卷积神经网络模型来模拟小麦所需的土壤环境，然后再用中国小麦产量数据对模型进行训练评估，结果表明这两个深度学习模型能够很好的预测粮食产量 (Wang, 2020)。虽然遥感技术能够获得有效的影响因素，但由于所需设备的昂贵性，在农民中不太使用，于是无人机技术获取影响因素在农产量预测中呈现递增的趋势。Nevavuori P 等人 (2019) 使用无人机获取当地的 NDVI 和 RGB 数据的作物产量，然后应用 CNN 对这些数据进行建模预测，并从 CNN 不同方面进行测试，结果表明卷积神经网络在 RGB 数据上的表现优于 NDVI 数据。Khaki (2020) 为了测试自己提出的新模型——CNN-RNN 对于农产量预测的稳定性，Khaki 学者通过对不同的模型的预测效果进行比较，比如使用 DFNN、RF 以及 LASSO 模型对美国玉米与大豆产量进行预测，然后再与 CNN-RNN 的预测效果进行比较，结果表明 CNN-RNN 模型预测误差显著小于其他几种预测模型。

### 1.2.2 国内研究现状

国内研究人员关于粮食产量预测方法主要采用传统的统计模型和时间序列分析方法。这些方法基于历史数据的趋势和周期性来进行预测，例如杨铁军 (2015) 使用修正的 ARIMA 模型对我国 1949-2011 年粮食产量数据进行分析，仿真验证模型可行性。然后吴越 (2020) 为了比较 ARIMA 模型以及 Holt 两参数指

数平滑法模型在不同预测年份的预测效果，使用长三角地区的粮食产量数据集，通过建立 ARIMA 模型以及 Holt 两参数指数平滑法模型来预测该地区粮食产量，实验结果表示，在不同预测年份时，模型预测的效果不同，ARIMA 模型对长预测年限的预测效果最好。张耀兰（2016）使用 GM（1，1）模型对安徽省 2006-2014 年的粮食产量进行分析，并预测安徽省未来粮食产量，得出安徽省产量呈现上升趋势。虽然 GM 模型在粮食产量预测中具有很好的适用性，但是在预测精度上然后有一定的不足（张模蕴，2020），于是张学者对非线性灰色模型进行改进，将灰色关联分析应用于 GM 模型中构建了一种新模型（OGM），然后使用该模型对粮食产量进行预测，并通过对比 GM 模型预测结果，得出 OGM(1,N)模型在粮食产量上的预测效果更好。粮食产量受多种因素影响，因此段霞学者（2021）为了预估未来山东省粮食产量，首先便对粮食产量的影响因素进行了回归分析，然后在回归分析的基础上建立了多元回归模型，最后验证了模型的拟合效果与预测效果都整体较好。

近年来，随着机器学习技术的快速发展，越来越多的研究人员开始探索将机器学习方法应用于粮食产量预测领域。例如，韩飞与高心怡（2020）对我国粮食产量进行预测，他们将粒子群算法和人工鱼群算法进行改进，然后使用我国粮食产量数据进行验证，实验结果验证了模型的可靠性。同年，李环等学者（2020），将粮食产量影响因素与粮食产量进行相关性分析，筛选出了 7 种与粮食产量极强相关的影响因素，并构建成预测数据集。然后分别建立多元线性回归模型、非线性自回归神经网络模型，并对我国粮食产量进行预测，最后对模型的预测结果进行对比分析。上述方法在粮食产量预测领域中都取得了一定的成果，能够有效的预测粮食产量，然而随着研究的深入，研究人员发现单一的模型在预测方面仍然存在一定的局限性，于是国内研究人员开始向组合模型方向进行研究。杨卫明等（2017）在预测我国粮食产量时，使用了多元线性回归模型与 DGM（1，1）模型相结合的组合模型，然后运用该模型对我国粮食产量进行预测，结果表明，组合模型的预测效果显著强于多元线性回归模型与 DGM（1，1）模型单独对粮食产量预测的效果。在机器学习方面，各种机器学习方法在粮食产量上预测的模型优势也各不一样，因此，邢聪仁学者（2019）将支持向量回归、BP 神经网络以

及 RFR 进行组合,并用该组合模型对安徽省粮食产量进行预测,预测结果表明组合模型预测的实验误差显著降低。王少航学者(2019)在对河南省粮食总产量进行预测时,首先使用了关联性分析筛选出 4 个影响因素,然后与粮食总产量构建为粮食产量数据集。使用该数据集建立 GM(0,4)模型、马尔科夫链优化后的灰色—马尔科夫模型以及加入遗传优化算法的灰色—马尔科夫模型。并用这些模型来预测河南省粮食产量,结果表明加入遗传算法后的组合模型预测误差显著性降低。郭亚菲学者(2019)在对传统单一的粮食产量预测模型进行研究后发现,单一模型在预测中存在数据遗落以及误差太大等问题。于是她采用小波分解对原始数据进行分解,接着将分解后的数据输入这些单一模型中,然后采用平均方法求的误差,结果发现使用这种组合的模型进行预测,模型拟合误差显著小于这两种单一的模型。胡程磊等学者(2021)在对江苏省粮食产量进行预测时,首先使用了 BP 神经网络进行预测,然后在 BP 模型的基础上加入了优化后的微粒群算法进行粮食产量预测。实验结果表明加入了优化后的微粒群算法的 BP 神经网络在预测的稳定性以及预测的精度上都有显著性提升。贾梦琦等(2021)对河北省粮食产量进行预测时,她将 ARIMA 模型,GRNN 模型以及 LSTM 模型分别对河北省粮食产量进行预测,然后再将这三个模型两两组合构建新的组合模型对粮食产量进行预测。实验结果表明,使用组合策略的模型在预测粮食产量上的预测效果更好,通过这项实验也为粮食产量预测提供了一种新思路。

### 1.2.3 文献述评

上述研究尽管已取得了富有建设性的成果,但仍然存在一定的拓展空间。在粮食产量预测相关实证研究中,我国研究人员在各自的研究领域中都取得了一定的成果。然后,对这些文献进行梳理总结后发现也存在一定的改进空间。传统的线性回归模型可能无法捕捉到粮食产量与各种影响因素之间的复杂非线性关系,导致预测准确度较低,决策树模型在处理高维数据和复杂数据关系时可能过拟合,导致模型泛化能力不足,预测结果不稳定。ARIMA、GM(1,1)等模型主要侧重于粮食产量的内在趋势影响,忽略了外在因素对粮食产量的影响。多元线性回归模型则考虑了外在因素对粮食产量的影响,但有时为了避免存在多重共线性的

存在,会造成大量信息的损失,并且多重共线性的存在会造成模型的误差存在,同时粮食产量预测是一个复杂的非线性系统,多元线性回归模型对非线性系统拟合较为复杂,而对于已有的神经网络模型,由于数据收集条件的限制,缺乏对粮食产量影响因素较为全面的概括,且模型选取数据量较小,造成输出结果精度不足。同时,仅仅依赖单一时间序列数据,未能全面考虑到实际农业生产中粮食产量受多种农业相关因素综合影响的情况。而国外研究人员在粮食产量的预测中主要是将机器学习与卫星遥感技术、气象产量预报技术进行结合,其核心的方法还是影响因素方法。然而,这些传统的影响因素方法对数据要求过高,而且获取困难,在预测过程中也容易受外部信息干扰,难以准确和可靠的预测粮食产量。而且主要侧重在太阳辐射水平以及土壤湿度上。因此,基于以上不足,在粮食产量预测中有必要进行相关的研究。由于随机森林能够处理大量的输入特征,并且在处理非线性关系和高维数据时表现良好。对于粮食产量预测这样的复杂问题,随机森林可以提供稳健的预测能力。BP神经网络具有强大的非线性建模能力。它适用于各种复杂的预测任务,并且在处理时间序列数据时具有一定的优势。对于粮食产量预测这样的时间相关性较强的问题,BP神经网络可以捕捉到数据中的复杂模式。同时支持向量回归能够处理高维数据和非线性关系。对于粮食产量预测这样的复杂问题,SVR可以有效地处理高维数据和非线性关系。再者,LSTM具有良好的记忆能力和长期依赖建模能力。它能够捕捉到数据中的长期依赖关系,并且在处理时间序列数据时表现优秀。对于粮食产量预测这样的时间序列预测问题,LSTM可以有效地捕捉到数据中的时间相关性和趋势。故本文采用相关算法进行研究并在预测过程中,综合权衡基于机器学习算法的各个模型的优势与限制,并根据实际情况选择最适合的模型,以确保获得更准确、稳定的预测结果。

## 1.4 研究框架

本文在对粮食产量预测时,首先进行了相关文献分析,明确了与粮食产量相关的影响因素,然后使用相关性分析方法筛选出与粮食产量极强相关的影响因素。接着使用RF、BP、SVR以及LSTM模型对粮食产量及其影响因素数据集进行建模预测,初步评估各个机器学习模型的预测效果。在粮食产量预测中单个模型进

行预测可能具有一定的随机性, 通过将多个单一模型的预测结果进行组合, 可以降低单一模型的偏差和方差, 从而提高整体预测的准确性。这些组合模型以往采用的是一种平均法策略。本文提出采用 Stacking 集成算法, 将 RF、BP、SVR 以及 LSTM 模型进行融合, 构建 Stacking 集成模型。Stacking 集成算法旨在通过结合多个基础模型的预测结果, 来构建一个更强大的元模型, 以提高整体预测的准确性和稳定性。为了比较 Stacking 集成模型的预测效果, 继续建立了平均组合模型进行对比分析。接着使用不同的数据集对模型进行训练来评估模型在面对不同数据特征和分布时的稳健性和可靠性。最后再用交叉验证方法对各个模型进行评估与分析。

本文共分为6章。第1章为选题背景、研究意义、文献综述; 第2章为相关概念与基础理论; 第3章、第4章和第5章为本文的核心部分, 主要研究粮食产量的单一模型实现, 基于Stacking集成算法的粮食产量预测以及粮食产量预测模型的对比分析; 第6章为本研究的总结与展望。具体分析框架如下:

第一章, 阐述本文的研究背景, 说明本研究的理论和现实意义。梳理国外相关文献, 简要评述国内外研究状况, 阐明研究思路和框架。

第二章, 对粮食产量的的基本概念以及影响因素进行介绍, 接着对本文所用的影响因素进行筛选, 并对本文所依据的的相关理论进行了简要的阐述。

第三章, 运用 LSTM、SVR、BP 以及 RF 模型来对我国粮食产量进行预测分析, 然后使用模型评估指标对模型进行初步评估。

第四章, 对第三章中的模型使用 Stacking 集成算法进行模型融合, 接着再使用平均集成思想对模型进行组合, 然后用这两种不同的算法构建的组合模型对我国粮食产量进行预测研究。

第五章, 分析各个模型在粮食产量上的预测效果。为了进一步评估模型在面对不同数据特征和分布时的稳健性和可靠性, 接着使用省域数据集对模型进行训练。然后再使用交叉验证的方法对第三章、第四章中构建的模型进行交叉验证, 再进一步评估模型的预测效果。

第六章, 总结研究所得的结论, 并对本文使用 Stacking 集成模型对粮食产量进行预测时的一些不足情况进行列举。

## 1.5 论文创新点

基于 Stacking 模型进行粮食产量预测的论文创新点主要有：将多个基础模型（如 BP 神经网络、随机森林、支持向量机以及 LSTM 模型）进行组合，形成三种不同的 Stacking 模型。通过优化 Stacking 模型的结构和参数，提高了对粮食产量预测的准确性和稳定性。并使用了省域数据集对模型的泛化性进行验证，最后使用交叉验证的方式进一步评估了模型预测的准确性和稳定性。

## 2 相关概念和理论基础

本章对粮食产量的基本概念以及影响因素进行介绍,接着对本文所用的影响因素进行筛选,并对本文所依据的相关理论进行了简要的阐述。

### 2.1 粮食产量预测的相关理论

#### 2.1.1 粮食产量的相关概念

##### (1)粮食产量

粮食产量是指在特定地区和特定时间内,生产的各种谷物和粮食作物的总产量。这些作物包括但不限于小麦、大米、玉米、大豆、稻谷、小米、高粱等主要用于人类食用的粮食作物。粮食产量通常以重量单位(如吨或千克)来表示,用于衡量农业生产的丰收程度以及满足人们的食物需求。粮食产量计算是指通过统计和分析农田实际收获的粮食数量以及相关数据,来确定某一特定地区、时间段内的农作物产量水平的过程。首先需要确定种植该作物的土地面积,可以通过土地调查、卫星遥感等方式获取相关数据,也可以通过农户调查来获得信息。然后,在农作物成熟季节,对种植地区进行收获调查,统计实际收获的粮食数量。这可以通过农民报告、田间调查、农业机械收割数据等方式进行。最后,根据实际收获的粮食数量和相应的种植面积,计算出单位面积上的粮食产量。产量计算可以采用简单的数学运算,如总收获量除以种植面积,也可以结合更复杂的统计方法和模型进行。

##### (2)粮食产量预测

粮食产量预测是指利用历史数据、气象信息、土壤状况、种植面积、农业技术等因素,通过数学统计方法或机器学习算法等手段,对未来一定时期内的粮食产量进行估计和预测的过程。这种预测通常用于农业规划、粮食市场供需分析、政府政策制定以及食品安全等领域。粮食产量预测的目标是为决策者提供有关未来粮食供应的信息,以便他们能够做出相应的调整和决策。预测粮食产量涉及到多个因素的复杂交互作用,因此需要综合考虑气候变化、农业技术进步、土地利用情况、政策影响等各种因素,以提高预测的准确性和可靠性。

## 2.1.2 粮食产量的影响因素

粮食产量的影响因素是指影响农作物生长、发育和最终产量的各种因素，这些因素可以分为自然因素和人为因素两大类。自然因素：自然因素包括气候、土壤、地形等。气候因素如降水量、温度、光照等直接影响着植物的生长季节和生长速率。土壤因素如土壤质地、肥力、水分含量等影响着植物的根系生长和养分吸收情况。地形因素如海拔高度、坡度等也会影响气候和水文条件，进而影响作物的生长环境。人为因素：人为因素包括农业技术、耕作管理、农业政策等。农业技术的进步、种植品种的选择、施肥、灌溉、病虫害防治等都会对产量产生直接影响。农业政策、市场需求和价格变动等也会影响农民的种植行为和农作物的产量水平。虽然通过分析影响因素对粮食产量的作用关系，建立影响因素与粮食产量的映射关联，可以实现粮食产量的较高精度预测，但是粮食产量的影响因素过多，建立的模型过于复杂、信息冗余也会导致预测模型精度下降。因此在建立预测模型前，本文采用相关性分析方法来探究粮食产量影响因素对粮食产量的作用关系，然后剔除一些弱相关性的影响因素。通过对我国 1949 年至 2022 年的粮食产量与其影响因素多变量时间序列，使用 Pearson 相关系数分析影响因素和粮食产量的相关性，对得到的相关系数数据结果如表 2.1 所示。

**表 2.1 粮食产量各个影响因素的相关性系数**

影响因素	X1	X2	X3	X4	X5	X6
相关系数	0.973	0.953	0.924	0.922	0.919	0.784
影响因素	X7	X8	X9	X10	X11	X12
相关系数	0.758	0.721	0.679	-0.081	-0.089	-0.911

其中，X1 表示有效灌溉面积，X2 表示农用化肥施用量，X3 表示粮食播种面积，X4 表示农业机械总动力，X5 表示农药使用量，X6 表示粮食除涝面积，X7 表示农林牧渔劳动力，X8 表示已建成水库数，X9 表示塑料膜使用量，X10 表示农产品生产价格指数，X11 表示生产资料指数，X12 表示粮食受灾面积。

相关性分析是一种可以用来研究两个或两个以上时间序列关联程度的统计学分析方法。两个变量之间是否存在关联或者关联强弱，可以使用皮尔逊相关系数作为相关程度的衡量标准，相关系数分布在 $[-1,1]$ 之间，相关系数大于 0，则表示序列之间正相关，小于 0 则表示负相关，等于 0 则表示不相关。相关系数的绝

对值越大，表示相关程度越大。当相关系数的绝对值大于 0.8 时，序列之间有极强的相关性。因此，本章筛选出相关系数绝对值大于 0.8 的影响因素来构建数据集。筛选出来的粮食产量影响因素的定义如下：

1.粮食作物有效灌溉面积是指实际有效利用灌溉水资源的土地面积。这包括通过灌溉设施（如灌溉渠道、滴灌系统、喷灌系统等）向农田供水的土地面积。有效灌溉面积的计算通常考虑了灌溉水的供应情况、土地的排水状况以及农作物对水资源的需求等因素。该指标对于评估农业水资源利用效率以及农田灌溉效果具有重要意义。

2.粮食播种面积是指在一定时期内，用于种植粮食作物的总面积。这包括了农民或农场主在耕种季节内种植的全部粮食作物的种植面积，如小麦、水稻、玉米等。粮食播种面积是农业统计中一个重要的指标，对于了解粮食生产潜力、制定农业政策、预测粮食产量等都具有重要意义。粮食播种面积的确定通常需要依赖于农业调查和统计数据，可以通过农户报告、农业部门的调查、卫星遥感等方式进行测算和统计。这一数据不仅反映了农业生产的规模和水平，还能够为农产品市场供需分析、粮食安全评估等提供重要参考依据。

3.农用化肥施用量指的是在农业生产中，农民向农田施用的化学肥料的总量。这包括各种类型的化肥，如氮肥、磷肥、钾肥等，以及复合肥料。农用化肥施用量通常以重量单位（如千克或吨）表示，用来衡量农业生产中对土壤提供营养元素的数量。

4.农药使用量指的是在农业生产中，用于防治病虫害、杂草等的农药的总量。这包括各种类型的杀虫剂、杀菌剂、除草剂等农药的施用量。农药使用量通常以重量单位（如千克或吨）或体积单位（如升或吨）表示。农药的使用对于提高农作物产量、改善农产品品质以及保护农作物免受病虫害侵害具有重要作用，但也需要谨慎使用以防止对环境和人类健康造成不良影响。因此，监测和控制农药使用量是农业管理和环境保护的重要任务之一。

5.农机总动力是指在生产过程中所用的所有农业机械设备的总功率或总驱动能力。这包括拖拉机、收割机、播种机、喷灌设备等各种农用机械设备的动力总和。通常以千瓦（kW）或马力（HP）为单位来表示。农业机械总动力是衡量

农业生产现代化程度和机械化水平的重要指标之一，它直接影响着农业生产效率和农产品的质量与产量。

6.粮食受灾面积是指在特定地区，由于自然灾害或人为因素导致的各种灾害（如洪涝、旱灾、病虫害等）而受到影响的种植面积。这包括粮食作物（如小麦、大米、玉米等）受到灾害影响的种植区域的总面积。通常，粮食受灾面积是在农业调查和统计中进行测量和记录的重要指标之一，用于评估灾害对农业生产的影响程度，以及制定相应的救灾和恢复计划。

这些与粮食产量相关的变量在粮食产量预测中具有重要作用。

## 2.2 机器学习相关算法基本原理

### 2.2.1 BP 神经网络

BP 神经网络是一种基于误差反向传播算法的人工神经网络模型，也称为多层感知器（Multilayer Perceptron）。它由输入层、隐藏层和输出层组成，具有前馈和反馈两个阶段。在前馈阶段，输入数据从输入层传递到输出层，不断通过隐藏层的神经元进行加权和激活。每个神经元都将输入的信号加权求和，然后通过激活函数进行非线性变换，产生输出信号传递给下一层。在反馈阶段，BP 神经网络使用误差反向传播算法（Backpropagation，简称 BP）来调整神经元之间的连接权重，以最小化输出与实际值之间的误差。具体来说，通过计算输出误差，并沿着网络反向传播这个误差，从而调整各层之间的连接权重，以减小误差。

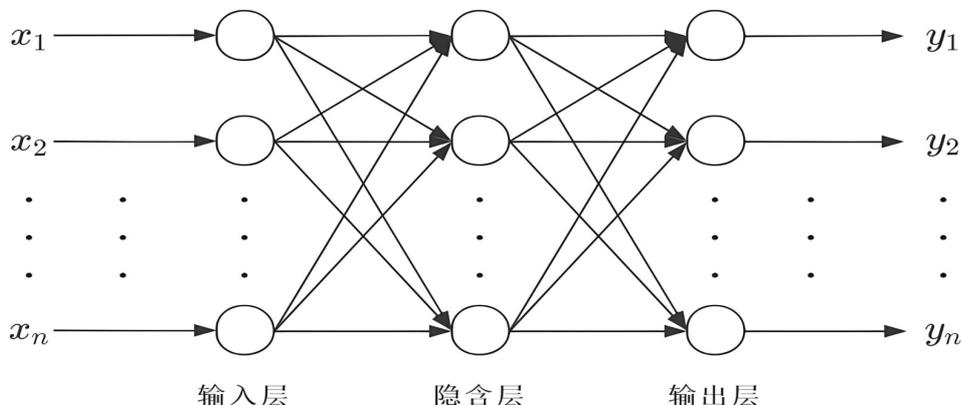


图 2.1 BP 神经网络结构

BP 神经网络的训练过程通常包括以下几个步骤：首先，随机初始化网络的连接权重和偏置。然后将输入数据传递到网络中，计算每一层的输出。接着计算输出层的误差，通常使用均方误差或交叉熵等损失函数来衡量预测值与实际值之间的差异。并且从输出层开始，将误差沿着网络反向传播，根据误差调整连接权重和偏置。再根据反向传播计算得到的梯度，使用优化算法（如随机梯度下降）更新网络参数，以减小误差。最后进行反复执行前向传播和反向传播步骤，直到达到停止条件（如达到最大迭代次数或误差足够小）。BP 神经网络算法流程如图 2.5 所示。

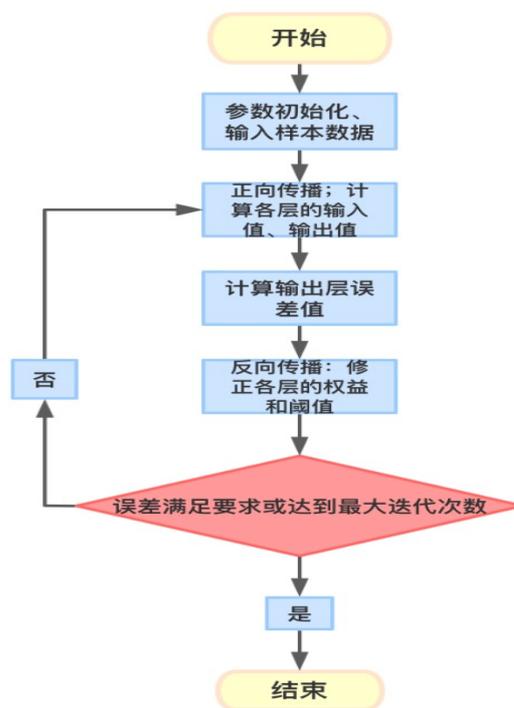


图 2.2 BP 神经网络算法流程图

### 2.2.2 支持向量回归

支持向量回归是一种基于支持向量机（Support Vector Machine）的回归算法。与传统的线性回归不同，SVR 旨在寻找一个非线性的函数来建模输入和输出之间的关系，并进行回归预测。SVR 的基本思想是通过将输入空间映射到高维特征空间，然后在特征空间中找到一个最优的超平面，使得该超平面与输入数据之间的间隔最大化，同时允许一定的误差范围。在回归任务中，SVR 的目标是最小

化预测值与实际观测值之间的误差,并且在保证预测误差不超过一定容忍度的情况下,尽量使间隔最大化。SVR 与 SVM 类似,它们都依赖于核函数来实现非线性映射。常用的核函数包括线性核、多项式核、径向基核(RBF 核)等,可以根据数据的特点和任务的需求选择合适的核函数。总的来说,SVR 适用于处理具有复杂非线性关系的回归问题,具有较好的泛化能力和抗噪声能力,在机器学习和数据建模领域得到了广泛的应用。

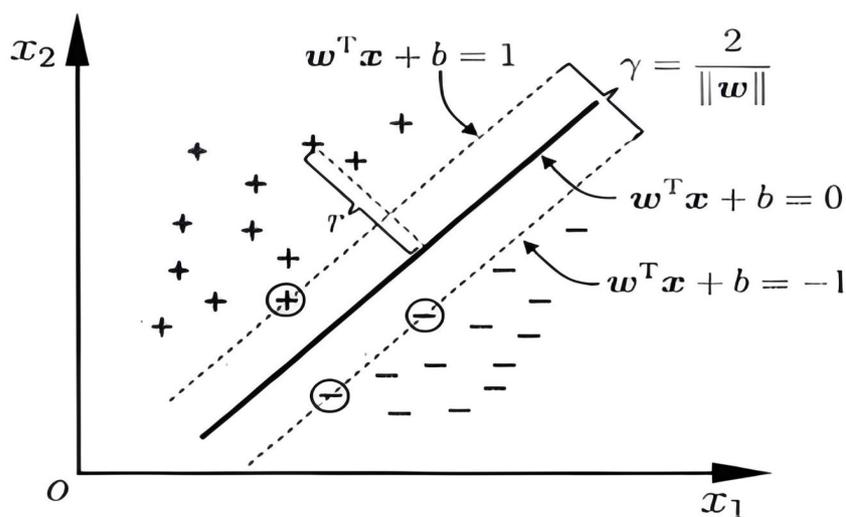


图 2.3 二维支持向量机

支持向量回归(SVR)在粮食产量预测中具有多个优点,SVR能够通过核函数实现非线性映射,可以捕捉粮食产量与各种影响因素之间复杂的非线性关系。这使得SVR能够更准确地建模和预测粮食产量,尤其是在存在非线性关系的情况下。而且SVR通过间隔最大化的方式确定超平面,因此对于训练数据中的噪声和异常值具有一定的鲁棒性。这使得SVR在处理真实世界中的农业数据时能够更好地应对数据的不确定性和复杂性。再加上SVR中的核函数可以根据具体问题和数据特征进行选择 and 调整,包括线性核、多项式核、径向基核等。这种灵活性使得SVR能够适应不同类型的数据和问题,提高了模型的适用性。

## 2.2.4 随机森林

随机森林是通过构建多个决策树来完成分类或回归任务。它由多个决策树组成,每个决策树都是一个分类器,且每个决策树的输出都是最终结果的投票。随

机森林的核心思想是通过集成多个决策树的预测结果来提高整体模型的性能和泛化能力。在构建随机森林时，首先会随机选择样本集和特征集，然后利用这些样本和特征来训练多个决策树。在每个决策树的构建过程中，会采用随机选择的样本和特征来进行节点的划分，以降低模型的方差和提高泛化能力。最后，对于分类任务，随机森林通过投票方式选择出现次数最多的类别作为最终预测结果；对于回归任务，则取多个决策树的预测值的平均值作为最终预测结果。随机森林具有很多优点，包括对于高维数据和大规模数据的处理能力强，具有较好的鲁棒性和泛化能力，并且能够有效地处理缺失值和异常值。此外，随机森林还可以评估特征的重要性，帮助理解数据和模型。因此，随机森林在实际应用中被广泛使用。而粮食产量受到多种因素的影响，如气候、土壤、种植方式等，这些因素之间可能存在复杂的非线性关系。随机森林作为一种集成学习方法，由多个决策树组成，能够很好地捕捉这些复杂的非线性关系，从而提高了预测的准确性。

### 2.2.1 长短期记忆神经网络 LSTM

#### (1) RNN

循环神经网络（RNN）是一种主要用于处理序列数据的神经网络结构。与传统的前馈神经网络不同，RNN 具有记忆性，能够保持对先前输入信息的记忆并将其应用于后续的输入数据中，因此在处理序列数据（如时间序列、自然语言、音频等）时具有很好的效果。RNN 的基本结构包括输入层、隐藏层和输出层。隐藏层中的神经元通过一个循环连接，可以将上一时刻的输出作为当前时刻的输入，形成了时间上的循环结构。这种设计使得 RNN 能够处理任意长度的序列数据，对于长度不固定的输入尤其有用。因此，使用 RNN 模型能够处理粮食产量这种时间序列问题，RNN 循环结构图如图 2.1 所示：

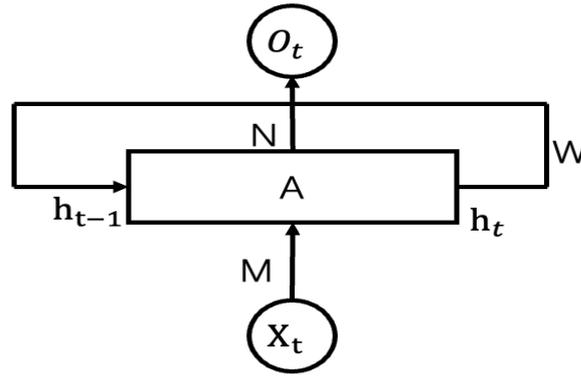


图 2.4 循环结构图

在这个结构图中， $X_t$ 表示输入序列的第 $t$ 个时间步， $h_t$ 表示隐藏状态（也称为记忆单元）在时间步 $t$ 的值， $O_t$ 表示输出序列的第 $t$ 个时间步。

RNN 的关键特点是隐藏状态的循环连接，它使得网络可以在处理序列数据时保留和传递信息。在每个时间步，输入 $X_t$ 和隐藏状态 $h_{t-1}$ 通过某个激活函数（如  $\tanh$ ）进行计算，得到新的隐藏状态 $h_t$ 。然后，根据隐藏状态 $h_t$ 计算输出 $O_t$ 。

## (2) LSTM

RNN 的训练过程通常使用反向传播算法。但是传统的 RNN 在面对较长的序列时，会出现梯度消失或爆炸的问题，导致难以有效训练长期依赖关系。为了解决这一问题，出现了一些改进型的 RNN 结构，如长短期记忆网络（LSTM）和门控循环单元（GRU）等，它们能够更好地捕捉长期依赖关系，并在很多序列数据处理任务中取得了较好的效果。

虽然 RNN 模型在理论上能够很好地对时间序列数据进行预测，但最终公式中存在的权重矩阵 $W^k$ 可能会导致梯度爆炸或者梯度消失现象，尤其在处理长序列数据时。此外，尽管我们使用了全局的数据，随着迭代次数的增加，传统的 RNN 模型最终只能记忆很短的全局信息，难以捕捉长期依赖关系。而长短期记忆神经网络（LSTM）能够有效地解决这些问题，使得模型能够更好地处理长序列数据并捕捉长期依赖关系。LSTM 内部包含四个关键的网络层，即输出门、输入门、遗忘门和记忆单元。这些门控单元负责控制信息的流动，决定哪些信息需要被记忆、遗忘或更新。通过这种机制，LSTM 能够更有效地处理长序列数据，并捕捉长期依赖关系，从而在处理时间序列数据时表现更为出色。长短期记忆神经网络结构图如图 2.5 所示：

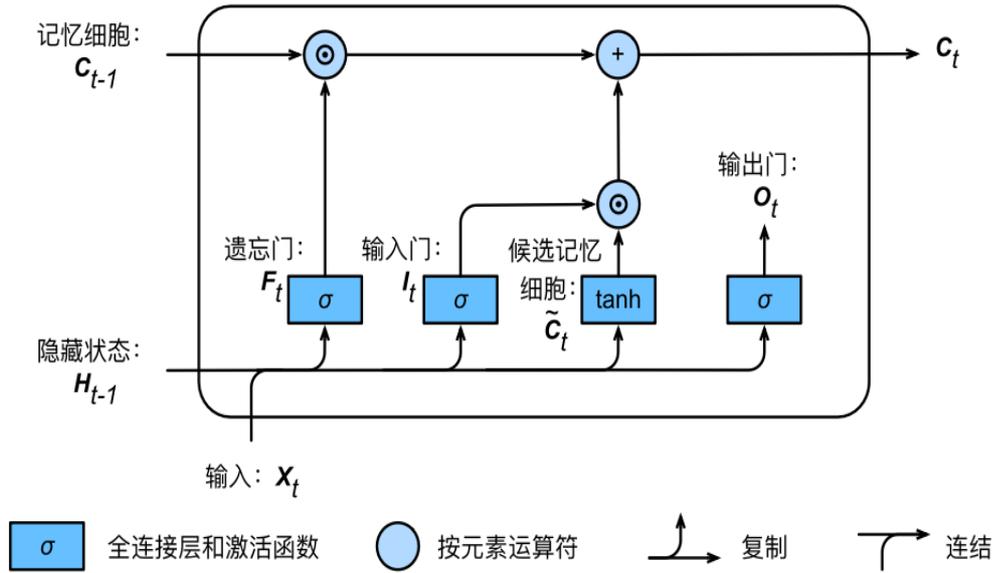


图 2.5 LSTM 循环结构图

由图可知，LSTM 模型的结构包括输入门：决定新输入信息的哪些部分将被加入到细胞状态 (Cell State) 中。遗忘门：决定细胞状态中的哪些信息将被遗忘，以适应新的输入。细胞状态：这是 LSTM 网络中的核心组件，用来传递信息和记忆状态。细胞状态可以通过输入门和遗忘门来调整和更新。输出门：决定细胞状态的哪些部分将被输出到当前时刻的隐藏状态和下一时刻的细胞状态中。

### 2.3 Stacking 算法与模型融合

Stacking 算法是一种模型融合技术，通过结合多个基本模型的预测结果来提高整体模型的性能。与简单的投票或平均法相比，Stacking 更进一步地利用了不同模型的预测结果，并通过训练一个元模型来组合这些预测结果。Stacking 算法的基本思想是将多个基本模型的预测结果作为新的特征，然后使用这些特征作为输入来训练一个元模型。这个元模型通常是一个简单的机器学习模型，如线性回归、逻辑回归等。元模型学习如何有效地组合基本模型的预测结果，以产生最终的整体预测结果。

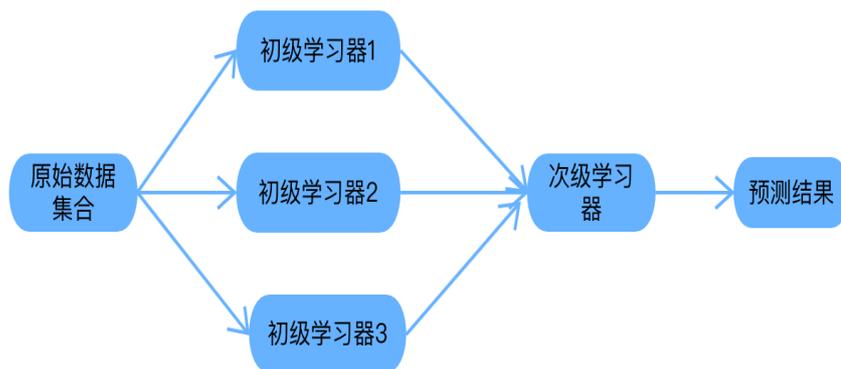


图 2.6 Stacking 算法的集成学习方式

图 2.6 为 Stacking 算法的基本流程图: 在使用 Stacking 算法时, 先将训练数据集分为训练集和验证集, 使用交叉验证技术来确保模型的泛化能力。这样可以更好地评估模型的性能, 并减少过拟合的风险。然后确保在第一层中使用的基本模型具有多样性, 即使用不同类型或参数设置的模型。这样可以增加模型的多样性, 提高集成模型的性能。其次在第二层模型中, 可以尝试不同的模型融合策略, 如加权平均、堆叠、投票等, 以找到最适合数据的组合方式。通过这种算法, 可以使 Stacking 算法更加有效地提高模型的准确度、降低误差, 并进一步扩展其应用范围。

## 2.4 交叉验证

交叉验证是一种统计学方法, 用于评估机器学习模型的性能和泛化能力。常用于训练模型时评估模型的预测能力, 以及选择模型的超参数。其基本思想是将原始数据划分为  $k$  个子集, 称为折叠 (folds), 然后进行  $k$  次模型训练和评估。在每一次训练中, 将其中一个折叠作为验证集, 剩下的  $k-1$  个折叠作为训练集。这样就可以获得  $k$  个模型及其对应的性能评估结果, 通常是准确度、均方误差等。最后, 将这  $k$  次评估结果的平均值作为模型的性能指标。交叉验证的优点是能够更准确地评估模型的性能, 因为每个样本都被用于验证过程; 更可靠地评估模型的泛化能力, 因为多次验证可以减少对单次验证结果的依赖性; 可以有效利用有限的的数据, 因为每个样本都能够被用于训练和验证。

常见的交叉验证方法包括 Holdout 验证、 $k$  折交叉验证等。

### (1) Holdout 验证

Holdout 验证是一种简单的交叉验证方法，通常用于评估模型的性能和泛化能力。它将原始数据集划分为两个互斥的部分：一个用于训练模型，另一个用于验证模型。Holdout 验证的优点包括简单易行、计算效率高，但也存在一些缺点，如验证集的划分可能会影响模型的评估结果，因此可能不够稳健。此外，当数据量较小时，验证集的大小可能会影响评估结果的可靠性。因此，在实际应用中，通常会采用更复杂的交叉验证方法，如 k 折交叉验证，以更准确地评估模型的性能和泛化能力。

### (2) K 折交叉验证

K 折交叉验证是一种常用的模型评估方法，用于评估机器学习模型的性能和泛化能力。在 K 折交叉验证中，数据被分成 K 个大小相似的互斥子集，称为折 (fold)。每次迭代中，其中的一个折被保留作为验证集，而剩余的 K-1 个折被用来训练模型。这个过程重复 K 次，每个折都有一次被作为验证集的机会。最后，K 次验证的结果被平均（通常是取平均值），以得到最终的性能指标。K 折交叉验证的主要优点是更准确地评估模型性能，减少了因数据划分不合理而引起的偏差；充分利用了数据集中的所有样本，避免了数据的浪费；对模型的性能进行了多次评估，结果更具有可靠性和稳定性。通常，K 的取值为 5 或 10，但也可以根据具体情况选择其他合适的数值。K 折交叉验证在模型选择、超参数调优和比较不同模型性能等方面被广泛应用。

## 2.5 本章小结

本章介绍了粮食产量的相关概念，如粮食产量的定义，影响因素的定义，并且使用了相关性分析来筛选与粮食产量极强相关的影响因素。接着对本文所用的影响因素进行筛选，并对本文所依据的机器学习理论、Stacking 集成算法概念、交叉验证思想进行了简要的阐述。

### 3 基于单一模型的粮食产量预测

本文使用了我国 1949 年至 2022 年的粮食产量以及相关的影响因素数据建立了单一的预测模型，在模型训练和测试过程中，对粮食产量的未来趋势进行预测和分析。

#### 3.1 基于 LSTM 模型的粮食产量预测分析

##### 3.1.1 数据来源及评价指标

本章选用了我国 1949 年至 2022 年的粮食产量以及相关的影响因素数据，数据来源于我国历年的统计年鉴。这些数据不仅能够帮助我们理解粮食产量的变化趋势，还能够揭示影响粮食产量的关键因素。因此，本文的研究结果具有一定的科学性和可信度，对于指导农业生产、制定政策以及粮食安全具有重要的意义。

在评估模型的拟合和预测效果时，我们通常使用各种评价指标来量化模型的性能。这些评价指标可以帮助我们了解模型在训练和测试数据上的表现，并根据需要做出调整或改进，这些常用的评价指标包括：均方根误差、平均绝对误差、平均绝对百分比误差。

这些评价指标能够帮助我们全面地评估模型的预测性能，从不同角度了解模型的准确度和稳健性。通常情况下，我们会根据具体的应用场景和数据特点选择合适的评价指标进行评估，并综合考虑多个指标来全面评价模型的表现。

##### 1. RMSE

$$RMSE = \sqrt{\frac{1}{n} \sum_{t=1}^N (s(t) - \hat{s}(t))^2}, \quad (3.1)$$

均方误差是预测值与真实值之间差异的平方的平均值，上述公式中的  $s(t)$  代表真实粮食产量， $\hat{s}(t)$  代表预估粮食产量。

##### 2. MAE

$$MAE = \frac{1}{N} \sum_{t=1}^N |s(t) - \hat{s}(t)|, \quad (3.2)$$

公式中的 $s(t)$ ,  $\hat{s}(t)$ 与公式 3.1 意思一致。

### 3. MAPE

$$MAPE = \frac{100\%}{N} \sum_{t=1}^N \left| \frac{\hat{s}(t) - s(t)}{s(t)} \right|, \quad (3.3)$$

公式中的 $s(t)$ ,  $\hat{s}(t)$ 与公式 3.1 意思一致, 平均绝对百分比误差 (Mean Absolute Percentage Error, MAPE) 是用于评估预测值与真实值之间差异的一种指标, 特别适用于回归问题。它表示平均预测误差在实际值的百分比中的大小。MAPE 的值越小, 表示模型的预测误差越小, 模型的预测性能越好。与均方根误差 (RMSE) 相比, MAPE 更直观地反映了预测误差的百分比大小, 因此在某些情况下更易于理解和解释。然而, MAPE 也有一些缺点, 例如当真实值为零或接近零时, 会导致分母为零而无法计算 MAPE。

### 3.1.2 LSTM 模型建模过程

#### 1. 数据标准化处理

本节所输入数据为粮食产量时间序列, 对粮食产量数据进行标准化处理。标准化处理公式如下:

$$y_i = \frac{x_i - x_{\min}}{x_{\max} - x_{\min}} \quad (3.4)$$

公式中的 $y_i$ ,  $x_i$ ,  $x_{\min}$ 、 $x_{\max}$ 分别表示归一化值、原始数据以及最小最大值。

#### 2.模型结构与参数调节

LSTM 模型在参数调节上主要是对时间步长调节: 时间步长决定了模型在处理序列数据时考虑的历史信息的数量。对于需要考虑长期依赖关系的问题, 适当增加时间步长可以帮助模型更好地捕捉到这种依赖关系, 从而提高模型的性能。然后对迭代次数调节: 迭代次数决定了模型训练的轮数, 更多的迭代次数通常会导致模型更充分地学习数据的特征和模式, 使模型收敛到更优的状态。接着进行批量调节: Batch-size 决定了每次更新模型参数时使用的样本数量。较大的 batch-size 通常会导致更稳定的梯度估计, 从而加快模型的收敛速度。最后进行 units 调节。

### 3.2.3 模型预测与结果分析

对模型参数进行调节之后接着我们使用调试好的 LSTM 模型对最近 1 年、3 年、5 年粮食产量进行预测，模型的预测值和实验误差如表 3.1、表 3.2、表 3.3 所示：

表 3.1 LSTM 模型预测 2022 年的预测产量及误差

年份	实际产量（万吨）	预测量（万吨）	RMSE	MAE	MAPE
2022	68653	69198	2543.28	2549.14	3.25%

表 3.2 LSTM 模型 2020-2022 年预测产量及误差

年份	实际产量（万吨）	预测量（万吨）	RMSE	MAE	MAPE
2020	66949	67320			
2021	68285	67551	2286.35	2677.35	3.36%
2022	68653	69834			

表 3.3 LSTM 模型 2018-2022 年预测产量及误差

年份	实际产量（万吨）	预测量（万吨）	RMSE	MAE	MAPE
2018	65789	66971			
2019	66384	67822			
2020	66949	66588	1486.84	1013.22	2.31%
2021	68285	67851			
2022	68653	68198			

从表 3.1、表 3.2、表 3.3 中的预测量以及误差指标可以看出，2022 年我国粮食实际产量为 68653 万吨，基于 LSTM 的粮食预测产量为 69834 万吨，LSTM 的粮食产量预测模型实验误差 RMSE 为 2543.28，MAE 为 2549.14，MAPE 为 3.25%。2020-2022 年 LSTM 的粮食产量预测模型实验误差 RMSE 为 2286.35，MAE 为 2677.35，MAPE 为 3.36%。预测 2018 年至 2022 年我国粮食总产量的预测模型实验误差 RMSE 为 1486.84，MAE 为 1013.22，MAPE 为 2.31%。

## 3.2 基于 BP 神经网络模型的粮食产量预测分析

本节使用的数据以及预处理与 3.1 节一致，故不在赘述。

### 3.2.1 BP 神经网络模型构建

使用 BP 神经网络模型对粮食产量及其影响因素数据集建模的主要步骤如下：首先对数据进行 Min-Max 归一化处理，该方法可以将数据线性地缩放到一个指定的范围内，通常是[0,1]或者[-1,1]。这种处理方法可以保留数据的分布信息，同时将数据映射到相同的尺度，有助于提高模型的训练速度和稳定性。由于本文使用 6 种粮食产量相关的影响因素作为特征变量构建粮食产量数据集，使用的特征变量较少，因此，采用 3 层网络结构构建 BP 神经网络。其次，隐藏层的节点数对神经网络的性能影响重大。在选择隐藏层节点数时，需要在充分学习数据特征的前提下尽可能降低模型的复杂度，以避免过拟合。通常情况下，隐藏层节点数不宜过多也不宜过少，需要根据数据集的规模和复杂度进行合理的选择。隐藏个数公式如下：

$$m = \sqrt{n + l} + a \quad (3.5)$$

确定隐藏层个数时，可以利用一个简单的经验公式来进行初步估计。该公式中，隐藏层的节点数（m）与输入层节点数（n）和输出层节点数（l）有关，同时考虑了一个调节常数（a），通常取值范围在[1,10]之间。在本文中我们构建的 BP 神经网络模型结构为 6-6-1，即包含一个隐藏层，每个隐藏层有 5 个节点，输出层有 1 个节点。作为激活函数，我们选择了 ReLU 函数。这一结构的设计旨在充分利用数据信息的同时，尽量避免模型过于复杂或简单，以达到较好的预测性能。

### 3.2.2 BP 神经网络模型训练与评估

接下来我们对 BP 神经网络进行训练，使用我国历年粮食产量数据集，并将测试数据集划分为 1 年、3 年、5 年。三种不同的年限来进行预测，预测结果数据如表 3.4、表 3.5、表 3.6 所示：

表 3.4 BP 神经网络模型预测 2022 年的预测产量及误差

年份	实际产量 (万吨)	模型预测量 (万吨)	RMSE	MAE	MAPE
2022	68653	67998	325.64	424.51	0.57%

表 3.5 BP 神经网络模型 2020-2022 年预测产量及误差

年份	实际产量 (万吨)	模型预测量 (万吨)	RMSE	MAE	MAPE
2020	66949	66588			
2021	68285	67851	983.85	1100.36	1.87%
2022	68653	68198			

表 3.6 BP 神经网络模型 2018-2022 年预测产量及误差

年份	实际产量 (万吨)	预测量 (万吨)	RMSE	MAE	MAPE
2018	65789	66150			
2019	66384	66982			
2020	66949	66588	1802.64	1998.18	2.81%
2021	68285	67851			
2022	68653	68198			

根据表 3.4、表 3.5 和表 3.6 的数据可知, BP 神经网络模型预测 1 年期的 MAPE 为 0.57%, 预测 3 年期的 MAPE 为 1.87%, 预测 5 年期的 MAPE 为 2.81%。所以该模型可以有效用于预测我国粮食产量, 并可提供对未来粮食生产情况的有用参考。

### 3.3 支持向量回归模型的粮食产量预测

支持向量机 (Support Vector Machine, SVM) 回归是一种机器学习算法, 用于解决回归问题。与传统的线性回归方法不同, SVM 回归通过使用核技巧在高维空间中进行非线性回归, 从而具有更强的泛化能力。在 SVM 回归中, 通过寻找最大间隔超平面来进行数据拟合, 同时允许一定程度的误差。支持向量指的是距离超平面最近的数据点, 它们决定了超平面的位置。SVM 在引入核函数后可以从低维度映射到更高维的空间上, 从而使得在原始空间中线性不可分的问题在特征空间中变为线性可分, 从而实现了非线性回归。所以 SVR 模型主要分为两步: 数据预处理、核函数选取。

#### (1) 数据预处理

支持向量回归的数据预处理步骤涵盖了数据清洗、特征处理、数据转换和划分等关键步骤，这些步骤能够为模型的训练和预测提供良好的基础。由于 SVR 模型对特征的尺度敏感，通常需要对特征进行缩放，使其尺度一致。常用的方法包括 Min-Max 缩放、Z-score 标准化等。

## (2) 核函数选取

选择合适的核函数对支持向量机回归的性能至关重要。核函数的选择取决于数据的特征以及问题的性质。SVR 模型的核函数主要有 Linear、Poly 和 Rbf。本节对不同的核函数依次替换试验，对比不同核函数在预测中的误差，选取预测误差最小的核函数进行预测。

表 3.7 核函数评价

核函数	惩罚系数 C	RMSE	MAE	MAPE
Linear	200	1878.31	1964.65	2.71%
Poly	50	2439.90	2142.61	3.11%
Rbf	1000	1323.14	1438.15	1.92%

通过对比不同核函数和惩罚系数在表 3.7 中的性能表现，本文选择了 RBF 核函数，并将惩罚系数 C 设定为 1000 来建立 SVR 模型，用于对粮食产量进行预测。经过模型训练和预测，得到了最近 1 年、3 年和 5 年的中国粮食产量预测结果如下：

表 3.8 SVR 模型预测 2022 年的预测产量及误差

年份	实际产量 (万吨)	模型预测量 (万吨)	RMSE	MAE	MAPE
2022	68653	68198	1365.39	1549.14	1.99%

表 3.9 SVR 模型 2020-2022 年预测产量及误差

年份	实际产量 (万吨)	预测量 (万吨)	RMSE	MAE	MAPE
2020	66949	66588			
2021	68285	67851	1250.63	1408.06	1.88%
2022	68653	68198			

表 3.10 SVR 模型 2018–2022 年预测产量及误差

年份	实际产量（万吨）	预测量（万吨）	RMSE	MAE	MAPE
2018	65789	65355			
2019	66384	66818			
2020	66949	66588	1302.64	1498.18	1.91%
2021	68285	67851			
2022	68653	68198			

根据表 3.8、表 3.9 和表 3.10 的数据，我们得到以下结论：

2022 年我国粮食实际产量为 68198 万吨，而基于 SVR 的粮食产量预测为 69834 万吨。SVR 模型的粮食产量预测误差如下：RMSE 为 1365.39，MAE 为 1549.14，MAPE 为 1.986%。2020 年至 2022 年的预测产量实验误差如表 3.9 所示，SVR 模型在这三年的预测中表现出了较好的准确性和稳定性。具体而言，2020 至 2022 年 SVR 的粮食产量预测模型实验误差为：RMSE 为 1250.63，MAE 为 1408.06，MAPE 为 1.88%。2018 年至 2022 年基于 SVR 的粮食产量预测模型实验误差如表 3.10 所示，SVR 模型在这五年的预测中也表现出了良好的性能。具体而言，2018 年至 2022 年 SVR 的粮食产量预测模型实验误差为：RMSE 为 1302.64，MAE 为 1498.18，MAPE 为 1.905%。这些实验结果显示，基于 SVR 的粮食产量预测模型在准确性和稳定性方面都表现出了较好的性能，为粮食产量预测提供了可靠的工具和参考。

### 3.4 基于随机森林模型的粮食产量预测分析

使用随机森林对粮食产量序列进行建模主要包括数据预处理和选择合适的决策树数量两个步骤：

#### （1）数据预处理

首先，将 1949 年至 2022 年的粮食产量数据划分为训练集和测试集。在训练集上进行序列，然后利用测试集进行预测。在对训练集进行预测时，需要选择合适的决策树数量。

#### （2）决策树数量的选取

决策树的数量过多会增加训练时间，可能导致模型过拟合；而数量过少则可能导致欠拟合。因此，需要选择合适数量的决策树。本节通过对不同决策树数量的实验，比较其对应的实验误差，选取误差最小的决策树数量作为最终模型的决策树数量，即  $n\_estimators=800$ 。随机森林模型的粮食产量预测误差如下：RMSE 为 1165.89，MAE 为 1149.63，MAPE 为 1.786%。

接下来，使用随机森林模型对 2020 年、2020-2022 年以及 2018 年至 2022 年的粮食产量进行预测。以下是预测的结果。

表 3.11 RFR 模型预测 2022 年的预测产量及误差

年份	实际产量（万吨）	模型预测量（万吨）	RMSE	MAE	MAPE
2022	68653	67978	1165.89	1149.63	1.79%

表 3.12 RFR 模型 2020-2022 年预测产量及误差

年份	实际产量（万吨）	模型预测量（万吨）	RMSE	MAE	MAPE
2020	66949	65963			
2021	68285	66734	794.93	934.67	1.49%
2022	68653	67278			

表 3.13 RFR 模型 2018-2022 年预测产量及误差

年份	实际产量（万吨）	预测量（万吨）	RMSE	MAE	MAPE
2018	65789	65455			
2019	66384	66618			
2020	66949	66252	625.64	898.18	0.93%
2021	68285	67755			
2022	68653	68247			

根据表 3.11、表 3.12 和表 3.13 的数据，我们可知：2022 年我国粮食实际产量为 68653 万吨，而使用随机森林模型预测的粮食产量为 67978 万吨。模型的实验误差如下：RMSE 为 1165.89，MAE 为 1149.63，MAPE 为 1.79%。对于 2020 年至 2022 年的预测粮食产量，模型的实验误差为：RMSE 为 794.93，MAE 为 934.67，MAPE 为 1.49%。对于 2018 年至 2022 年的预测粮食产量，模型的实验误差为：RMSE 为 625.64，MAE 为 898.18，MAPE 为 0.93%。这些结果表明，

随机森林模型在我国粮食总产量预测方面表现良好,具有较高的准确性和稳定性。该模型可用于对我国粮食产量进行可靠的预测,为决策和规划提供重要参考。

### 3.5 本章小结

在本章中,主要建立了 LSTM、BPNN、RFR 以及 SVR 模型对我国粮食产量进行了预测,并且根据模型评价指标对各个模型的拟合效果进行了对比分析。在对最近 1 年、3 年以及 5 年预测中,整体预测效果最优的依次是随机森林、BP 神经网络、SVR 最后是 LSTM 模型。

## 4 基于 Stacking 算法的粮食产量预测

前一章节我们构建了四种机器学习的方法对粮食产量进行预测，四种机器学习方法在预测效果上各不相同，接下来我们使用 Stacking 算法将这四种机器学习模型进行模型融合。Stacking 算法在粮食产量预测中的基本算法框架如下：

### 1.数据准备：

收集和整理历史粮食产量数据以及可能影响粮食产量的各种特征数据（粮食作物播种面积、有效灌溉面积、农用化肥施用量、农业机械总动力、农药使用量、粮食受灾面积）。将数据划分为训练集和测试集，用于模型训练和评估。

### 2.基学习器选择：

根据第三章模型评估的结果，我们首先选择其中两种预测效果最好模型作为基学习器预测模型，然后再依次加入预测效果较前两种效果一般的两种模型来增加模型的多样性和泛化能力。本文采用的基学习器为长短期记忆神经网络、BP 神经网络、随机森林、支持向量机。

### 3.训练基学习器：

使用训练集对每个选择的基学习器进行训练，得到各自的预测模型。

### 4.生成预测结果：

使用训练好的基学习器对测试集进行预测，得到每个模型的预测结果。

### 5.元学习器训练：

将基学习器的预测结果作为特征，将粮食产量作为目标，构建一个元学习器。使用训练集的基学习器预测结果作为输入，真实粮食产量作为输出，训练元模型。

### 6.预测：

使用训练好的元学习器对测试集的基学习器预测结果进行组合，得到最终的粮食产量预测结果。

### 7.评估和优化：

对预测结果进行评估，常用的评价指标包括均方根误差（RMSE）、平均绝对误差（MAE）等。根据评估结果对算法进行调优，可能包括选择更合适的基学习器、调整元学习器的参数等。

基于 Stacking 算法在粮食产量预测中的基本算法框架图如图 4.1 所示。

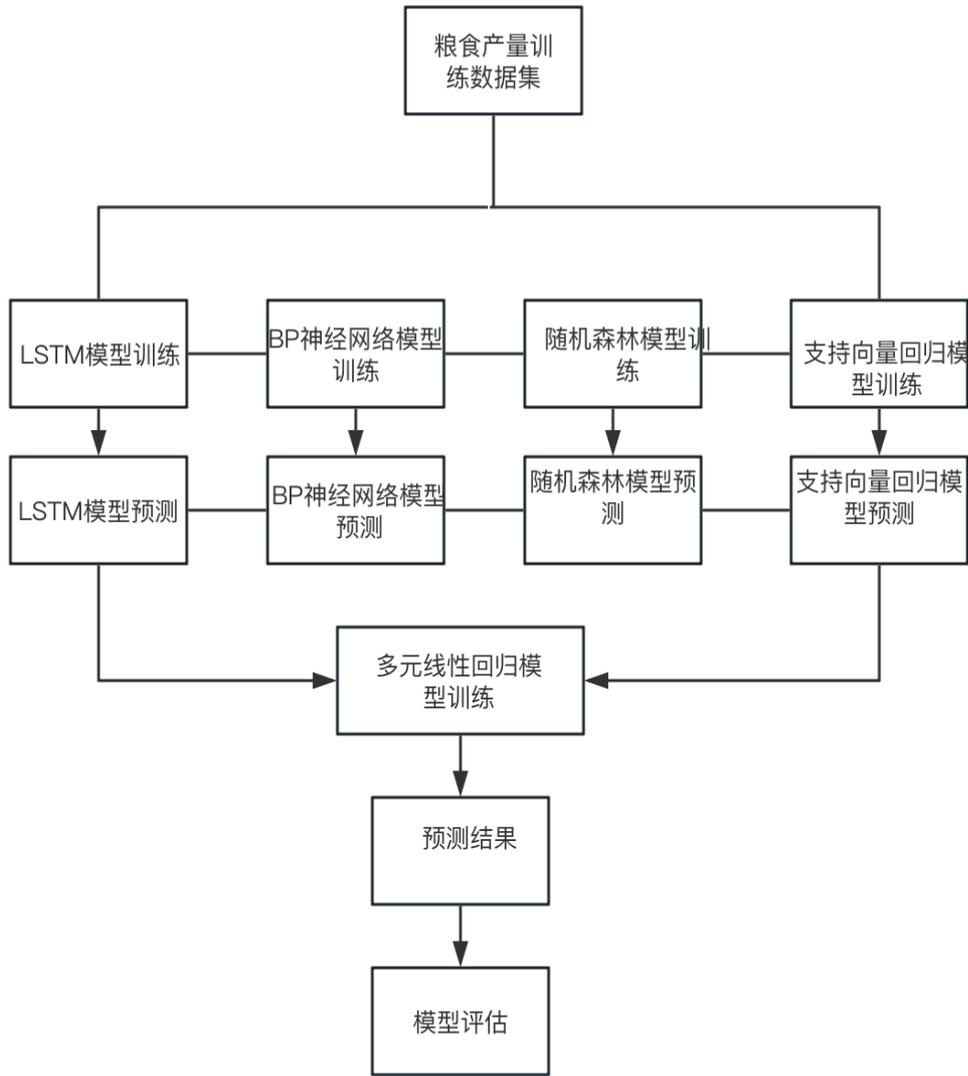


图 4.1 Stacking 算法在粮食产量预测中的基本算法框架图

#### 4.1 Stacking 集成模型构建

Stacking 集成算法是一种强大的机器学习技术，旨在提高预测模型的准确性和鲁棒性。该算法通过结合多个基础模型的预测结果，构建一个元模型来进行最终的预测。具体而言，Stacking 算法分为两个阶段：首先，在训练阶段，多个不同类型的基础模型被训练并用于生成预测结果；接着，在元模型训练阶段，这些基础模型的预测结果被用作输入数据，通过另一个模型（通常是线性回归模型）进行整合，从而得到最终的预测结果。通过利用多个模型的优势，Stacking 算法能够显著提高预测性能，尤其在处理复杂数据和挑战性任务时表现突出。

基本模型训练阶段：使用训练数据集来训练多个不同的基本模型，例如决策树、支持向量机、随机森林等。

元模型训练阶段：在该阶段，通过使用基本模型的预测结果来训练一个元模型（也称为次级模型），以提高性能。元模型可以是简单的线性模型，也可以是更复杂的模型，如支持向量机或者另一个集成模型。Stacking 的关键是使得基本模型和元模型之间的预测结果具有多样性，从而使得元模型能够有效地将这些结果整合为最终的预测。这通常通过使用不同的训练数据、不同的特征子集或者不同的算法来实现。

### (1) 初级学习器

使用 Stacking 算法首先在训练数据集来训练多个不同的基本模型，本文使用长短期记忆神经网络、BP 神经网络模型、随机森林和支持向量回归作为初级学习器，然后对我国最近 1、3、5 年的粮食产量进行预测时，在对预测效果进行初步评估后，然后构建 Stacking 集成模型，我们首先选择随机森林和 BP 神经网络模型作为第一层的基学习器，然后在第一层基学习器中依次加入支持向量回归、长短期记忆神经网络作为第一层的基学习器。构建三种不同的第一层 Stacking 算法的集成模型，然后在第二层中，我们需要对元学习器进行选择。

### (2) 元学习器的选择

在元学习器上，因为粮食产量与其相关影响因素变量数据集是一个回归问题，所以使用多元线性回归模型作为第二层的元学习器，然后构建了三种 Stacking 集成模型，即 RF-BPNN-Stacking、RF-BPNN-SVR-Stacking 以及 RF-BPNN-SVR-LSTM-Stacking 模型，这些模型可以充分利用各个单一模型的优势，并通过多元线性回归模型的组合来进一步提高粮食产量预测的准确性和稳定性。

## 4.2 Stacking 集成模型训练与评估

在构建好了模型后，我们需要对这几种模型进行训练与评估，为了能够与单一模型进行对比，本章模型训练数据集与第三章使用的数据集一致，并将数据集的测试数据划为 1 年期、3 年期以及 5 年期进行预测。

### (1) RF-BPNN-Stacking 模型

使用 RF-BPNN-Stacking 模型对我国最近 1 年、3 年、5 年的粮食产量进行预测，最终得到的预测数据如表 4.1 所示：

表 4.1 RF-BPNN-Stacking 模型预测我国粮食总产量的误差

预测年份	MAPE	MAE	RMSE
2020	1.33%	1097.83	1163.49
2020-2022	0.62%	517.53	583.19
2018-2022	1.53%	1290.60	1356.26

根据表 4.1 的结果，使用 RF-BPNN-Stacking 模型对我国最近 1 年、3 年、5 年的粮食产量进行预测时，预测 2022 年的我国粮食产量的 RMSE 为 1163.49，MAE 为 1097.83，MAPE 为 1.33%。预测 2020 年至 2022 年的我国粮食产量的 RMSE 为 583.19，MAE 为 517.53，MAPE 为 0.62%。预测 2018 年至 2022 年的我国粮食产量的 RMSE 为 1356.26，MAE 为 1290.60，MAPE 为 1.53%。这些结果表明，RF-BPNN-Stacking 模型取得了较好的拟合效果。该模型在最近几年的粮食产量预测中表现良好，其均方根误差、平均绝对误差和平均绝对百分比误差都较低，显示出了较高的预测准确性和稳定性。

#### (2) RF-BPNN-SVR-Stacking 模型

使用 RF-BPNN-SVR-Stacking 模型对我国最近 1 年、3 年、5 年的粮食产量进行预测，最终得到的预测数据如表 4.2 所示：

表 4.2 RF-BPNN-SVR-Stacking 模型预测我国粮食总产量的误差

预测年份	MAPE	MAE	RMSE
2020	0.75%	619.08	684.74
2020-2022	0.69%	492.70	358.36
2018-2022	1.86%	1951.64	2017.30

根据表 4.2 的结果，使用 RF-BPNN-SVR-Stacking 模型对我国最近 1 年、3 年、5 年的粮食产量进行预测，在预测 2022 年的我国粮食产量的 RMSE 为 684.74，MAE 为 619.08，MAPE 为 0.75%。预测 2020 年至 2022 年的我国粮食产量的 RMSE 为 358.36，MAE 为 492.70，MAPE 为 0.69%。预测 2018 年至 2022 年的

我国粮食产量的 RMSE 为 2017.30, MAE 为 1951.64, MAPE 为 1.86%。这些结果显示, RF-BPNN-SVR-Stacking 模型取得了较好的拟合效果。在预测最近几年的粮食产量时, 该模型表现良好, 其均方根误差、平均绝对误差和平均绝对百分比误差都较低, 显示出了较高的预测准确性和稳定性。此外, RF-BPNN-SVR-Stacking 模型在 1 年期预测的效果比 RF-BPNN-Stacking 集成模型预测效果更优。然而, 在预测最近 5 年的粮食产量时, RF-BPNN-SVR-Stacking 模型的拟合效果不如 RF-BPNN-Stacking 集成模型, 其预测指标均高于前者。

### (3) RF-BPNN-SVR-LSTM-Stacking 模型

使用 RF-BPNN-SVR-LSTM-Stacking 模型对我国最近 1 年、3 年、5 年的粮食产量进行预测的结果如表 4.3 所示。

表 4.3 RF-BPNN-SVR-LSTM-Stacking 预测我国粮食总产量的误差

预测年份	MAPE	MAE	RMSE
2020	0.79%	652.09	717.75
2020-2022	1.45%	610.10	675.76
2018-2022	1.17%	932.10	997.76

根据表 4.3 的结果, 使用 RF-BPNN-SVR-LSTM-Stacking 模型对我国最近 1 年、3 年、5 年的粮食产量进行预测, 在预测 2022 年的我国粮食产量的 RMSE 为 717.75, MAE 为 652.09, MAPE 为 0.79%。预测 2020 年至 2022 年的我国粮食产量的 RMSE 为 675.76, MAE 为 610.10, MAPE 为 1.45%。预测 2018 年至 2022 年的我国粮食产量的 RMSE 为 997.76, MAE 为 932.10, MAPE 为 1.17%。这些结果显示, RF-BPNN-SVR-LSTM-Stacking 模型取得了较好的拟合效果。在预测最近几年的粮食产量时, 该模型表现良好, 其均方根误差、平均绝对误差和平均绝对百分比误差都较低, 显示出了较高的预测准确性和稳定性。此外, 在对比 RF-BPNN-SVR-LSTM-Stacking 模型和 RF-BPNN-SVR-Stacking 模型的预测效果时, 发现在预测最近 1 年的粮食产量时, 两者的预测指标相当。然而, 在预测最近 3 年和 5 年的粮食产量时, 模型的拟合效果不如 RF-BPNN-SVR-Stacking 模型。

根据上述实验结果可知,使用 RF-BPNN-Stacking 模型、RF-BPNN-SVR-Stacking 模型、RF-BPNN-SVR-LSTM-Stacking 模型对我国最近 1 年、3 年、5 年的粮食产量进行预测,在预测 2022 年的我国粮食产量的 MAPE 分别为 1.33%、0.75%、0.79%。预测 2020 年至 2022 年的我国粮食产量的 MAPE 分别为 0.62%、0.69%、1.45%。预测 2018 年至 2022 年的我国粮食产量的 MAPE 为 1.53%、1.86%、1.17%。这些结果显示,这三个 Stacking 模型取得了较好的拟合效果。在预测 2022 年的我国粮食产量时 RF-BPNN-SVR-Stacking 模型预测效果最好;在预测 2020 年至 2022 年的我国粮食产量时 RF-BPNN-Stacking 模型预测效果最好;在预测 2018 年至 2022 年的我国粮食产量时 RF-BPNN-SVR-LSTM-Stacking 模型预测效果最好。

### 4.3 使用平均法的集成模型

为了更好的对比 Stacking 集成算法构建的几种模型,本节采用平均组合策略的方法讲前一节中的模型进行集成,得到三种平均集成模型,并使用相同数据集对构建的模型进行拟合。

#### (1) RF-BPNN-AVG 模型

使用 RF-BPNN-AVG 模型对我国最近 1 年、3 年、5 年的粮食产量进行预测,预测结果如表 4.4 所示。

表 4.4 RF-BPNN-AVG 模型预测我国粮食总产量的误差

预测年份	MAPE	MAE	RMSE
2020	1.00%	1460.11	1547.44
2020-2022	1.68%	1900.99	2215.22
2018-2022	1.19%	1659.34	1498.88

根据表 4.4 的结果,使用 RF-BPNN-AVG 模型对我国最近 1 年、3 年、5 年的粮食产量进行预测,在预测 2022 年的我国粮食产量的 RMSE 为 1547.44, MAE 为 1460.11, MAPE 为 1.00%。预测 2020 年至 2022 年的我国粮食产量的 RMSE 为 2215.22, MAE 为 1900.99, MAPE 为 1.68%。预测 2018 年至 2022 年的我国

粮食产量的 RMSE 为 1498.88, MAE 为 1659.34, MAPE 为 1.19%。这些结果显示, RF-BPNN-AVG 模型取得了较好的拟合效果。在预测最近几年的粮食产量时, 该模型表现良好, 其均方根误差、平均绝对误差和平均绝对百分比误差都较低, 显示出了较高的预测准确性和稳定性。

### (2) RF-BPNN-SVR-AVG 模型

使用 RF-BPNN-SVR-AVG 模型对我国最近 1 年、3 年、5 年的粮食产量进行预测, 预测结果如表 4.5 所示。

表 4.5 RF-BPNN-SVR-AVG 模型预测我国粮食总产量的误差

预测年份	MAPE	MAE	RMSE
2020	1.02%	1492.59	1603.09
2020-2022	0.41%	277.22	296.01
2018-2022	2.26%	2417.99	2465.69

由表 4.5 可知, 使用 RF-BPNN-SVR-AVG 模型对我国最近 1 年、3 年、5 年的粮食产量进行预测时, 预测 2022 年的我国粮食产量的 RMSE 为 1603.09, MAE 为 1492.59, MAPE 为 1.02%, 预测 2020 年至 2022 年的我国粮食产量的 RMSE 为 296.01, MAE 为 277.22, MAPE 为 0.41%, 预测 2018 年至 2022 年的我国粮食产量的 RMSE 为 2465.69, MAE 为 2417.99, MAPE 为 2.26%。RF-BPNN-SVR-AVG 模型与 RF-BPNN-AVG 集成模型在最近 1 年期的预测效果相当, 而在最近三年期的预测中, 优于 RF-BPNN-AVG 集成模型。在最近 5 年期预测效果却又不如 RF-BPNN-AVG 集成模型。

### (3) RF-BPNN-SVR-LSTM-AVG 模型

RF-BPNN-SVR-LSTM-AVG 模型预测结果数据如表 4.6 所示。

表 4.6 RF-BPNN-SVR-LSTM-AVG 模型预测我国粮食总产量的误差

预测年份	MAPE	MAE	RMSE
2020	1.94%	2427.05	2292.28
2020-2022	2.37%	2300.75	2413.44
2018-2022	2.34%	2828.07	2498.88

由表 4.6 可知,使用 RF-BPNN-SVR-LSTM-AVG 模型对我国最近 1 年、3 年、5 年的粮食产量进行预测时,预测 2022 年的我国粮食产量的 RMSE 为 2292.28, MAE 为 2427.05, MAPE 为 1.94%, 预测 2020 年至 2022 年的我国粮食产量的 RMSE 为 2413.44, MAE 为 2300.75, MAPE 为 2.37%, 预测 2018 年至 2022 年的我国粮食产量的 RMSE 为 2498.88, MAE 为 2828.07, MAPE 为 2.34%。

从这三个平均法集成模型预测结果来看,模型的整体效果良好, RF-BPNN-AVG 模型和 RF-BPNN-SVR-AVG 模型在预测 2022 年粮食总产量的预测效果较好,预测的平均绝对百分比误差为 1.00%和 1.02%。RF-BPNN-SVR-AVG 在预测 2020 年至 2022 年最近 3 年的预测效果优于其它模型,预测的平均绝对百分比误差为 0.41%。在最近 5 年预测效果中, RF-BPNN-AVG 模型的预测效果优于其它模型,预测的平均绝对百分比误差为 1.19%。

#### 4.4 本章小结

本章节首先介绍了 Stacking 集成算法的基本框架,接着使用该算法对四种机器学习模型进行融合,由于这 4 种机器学习模型在同一数据集上预测效果不一,所以我们先选择两种预测效果最后的机器学习模型构建第一种 Stacking 集成模型,接着依次加入其它两种机器学习模型构建不同的 Stacking 集成模型。并对构建的模型进行训练与评估,然后为了更好的进行对比分析,再接着构建平均法集成模型进行训练评估。实验结果表明,基于 Stacking 集成算法的集成模型在粮食总产量预测方面取得了较好的拟合效果,基于平均法集成模型在粮食预测上表现良好。但在不同年限上最优的预测模型不同,在预测一年期,最优模型为 RF-BPNN-SVR-Stacking 集成模型,在预测 3 年期的粮食产量数据集时, RF-BPNN-SVR-AVG 模型预测效果最好,在预测 5 年期的粮食产量数据集时 RF-BPNN-SVR-LSTM-Stacking 模型展现出更好的预测效果。

## 5 粮食产量预测模型的对比分析

本章将采用省域数据集作为对比数据集进行分析,然后再使用交叉验证的方法对模型进行验证。

### 5.1 预测模型的对比与分析

本文首先对四种单一的机器学习模型进行对比分析,然后在使用 Stacking 集成方法构建的三种模型与使用平均法构建的三种集成模型进行分析。通过比较这些模型的 MAPE 指标,可以评估它们的预测准确性和稳定性。较低的 MAPE 值意味着模型的预测更接近实际值,表明其具有更好的预测性能。

#### (1) 单一模型

使用四种单一的机器学习对粮食产量数据集进行建模预测,并构建不同年限的测试数据集进行测试,模型的评价指标如下表所示。

表 5.1 单一模型预测粮食产量的平均绝对误差

预测年份	BP 神经网络	随机森林	支持向量回归	长短期记忆神经网络
2022	424.51	1149.63	1549.14	2537.14
2020 至 2022	1100.36	934.67	1408.06	2677.35
2018 至 2022	1998.18	898.18	1473.17	1013.22

表 5.2 单一模型预测粮食产量的平均绝对百分比误差

预测年份	BP 神经网络	随机森林	支持向量回归	长短期记忆神经网络
2022	0.57%	1.79%	1.99%	3.25%
2020 至 2022	1.87%	1.49%	1.88%	3.36%
2018 至 2022	2.81%	0.93%	1.91%	2.31%

根据表 5.1 和表 5.2 的数据可知,对于 2022 年我国粮食产量的预测,长短期记忆神经网络的平均绝对误差和平均绝对百分比误差都比其他三种模型大,预测效果不如其他三种模型。而 BP 神经网络模型则表现出较小的平均绝对误差和平均绝对百分比误差,预测效果更好。在对 2020 年至 2022 年以及 2018 年至 2022 年的粮食总产量进行预测时,随机森林模型展现出更好的预测效果。这表明随机森林模型在长期和短期粮食产量预测方面都具有良好的适用性和准确性。所以,

针对不同时间段的粮食产量预测任务，不同的模型可能会表现出不同的优劣势。在选择合适的预测模型时，需要根据具体的预测需求和数据特点进行综合考虑。

## (2) Stacking 算法与平均法集成模型

使用 Stacking 算法与平均法集成模型预测不同年限的粮食产量时，模型的评价指标结果如如表 5.3、表 5.4 所示。

表 5.3 六种集成模型预测粮食总产量的平均绝对百分比误差

预测年份	RF-BPNN-Stacking	RF-BPNN-SVR-Stacking	RF-BPNN-SVR-LSTM-Stacking	RF-BPNN-AVG	RF-BPNN-SVR-AVG	RF-BPNN-SVR-LSTM-AVG
2022	1.33%	0.75%	0.79%	1.00%	2.62%	1.94%
2020-2022	0.62%	0.69%	1.45%	1.68%	0.41%	2.37%
2018-2022	1.53%	1.86%	1.17%	1.19%	2.56%	2.34%

表 5.4 六种集成模型预测粮食总产量的平均绝对误差

预测年份	RF-BPNN-Stacking	RF-BPNN-SVR-Stacking	RF-BPNN-SVR-LSTM-Stacking	RF-BPNN-AVG	RF-BPNN-SVR-AVG	RF-BPNN-SVR-LSTM-AVG
2022	1097.83	619.08	652.09	1460.11	1492.59	2427.05
2020-2022	517.53	492.70	610.10	1900.99	277.22	2300.75
2018-2022	1290.60	1951.64	932.10	1659.34	2417.99	2828.07

根据表 5.3 和表 5.4 的数据可知，两种不同算法的集成模型在对我国不同年限的粮食产量进行预测时，在不同年限上最优的预测模型不同，在预测一年期，最优模型为 RF-BPNN-SVR-Stacking 集成模型，在预测 3 年期的粮食产量数据集时，RF-BPNN-SVR-AVG 模型预测效果最好，在预测 5 年期的粮食产量数据集时 RF-BPNN-SVR-LSTM-Stacking 模型展现出更好的预测效果。综上所述，不同的集成模型在不同年份的粮食产量预测中表现出不同的优势。选择合适的预测模型需要根据具体的预测需求和数据特点进行综合考虑，以确保获得最准确的预测结果。

## 5.2 对比数据集预测结果

为了比较不同数据集下的模型训练和预测效果，我们将采用省级粮食产量数据集进行对比。本文使用江西省 1949 年至 2022 年的粮食产量数据以及影响因素作为对比数据集，进行模型的训练和测试。将数据集划分为训练数据集与测试集。测试数据集分别用 1 年期，3 年期、5 年期作为测试数据集。然后使用模型对测试集数据进行预测。这样的比较将有助于评估模型在不同数据集上的泛化能力和预测效果，从而更好地理解模型在不同场景下的表现。

在对比数据集中，各种单一机器学习预测的评价指标如表 5.5 中所示。

表 5.5 单一模型预测粮食产量的平均绝对百分比误差

预测年份	BP 神经网络	支持向量回归	随机森林	长短期记忆神经网络
2022	0.63%	2.19%	1.29%	3.19%
2020 至 2022	1.77%	1.13%	1.49%	4.03%
2018 至 2022	1.81%	1.71%	0.73%	3.31%

根据表 5.5 的数据，使用江西省 1949 年至 2022 年的粮食产量数据与影响因素的对比数据集进行模型训练与测试时，发现模型的预测效果与使用全国数据集时存在差异。具体来说：BP 神经网络模型在预测 2022 年的效果优于预测 2020 年至 2022 年和 2018 年至 2022 年的效果。这表明 BP 神经网络模型在特定省份数据集上的表现可能更好。长短期记忆神经网络模型的预测误差较大，表现不佳。这可能是由于该模型在处理省级数据集时遇到了一些挑战或者数据特征不够适合该模型的原因导致的。SVR 在预测最近 3 年的预测误差较小，而在预测 2022 年时的预测效果一般。这可能表明支持向量回归模型对于江西省的历史数据具有较好的泛化能力，但对于未来数据的预测表现一般。随机森林算法的预测效果较好，模型拟合情况良好。这表明随机森林算法在处理省级粮食产量数据集时能够取得良好的预测结果。所以不同模型在省级数据集上的表现可能会有所差异，需要根据具体情况选择合适的模型以获得更准确的预测结果。

## 5.2 使用交叉验证的预测模型对比与分析

交叉验证也即是对样本数据进行重复利用，它的基本原理是对预测数据集进行多次切分，然后把切分开来的数据集用作训练数据集和测试数据集。接着对这些数据集进行反复多次训练与测试。本节我们将采用交叉验证对构建的模型进行评估。交叉验证的形式有很多种，本文采用 K 折交叉验证，在 K 折交叉验证中，我们选取 5 折和 10 折交叉验证。5 折交叉验证的原理是把数据集分为 5 份，一份用作测试数据集，其它的 4 份用作训练数据集，依次反复，进行 5 次实验，并最终取 5 次实验的平均值作为最终结果。10 折交叉验证则是将数据集划分为 10 份，其中一份作为预测数据集，其它 9 份作为测试数据集，然后进行 10 次实验，最后取这十次实验的平均值作为最后的预测结果，最后将得到的评估结果进行汇总。通过使用交叉验证方式对模型进行评估，可以更客观地评估模型的性能，减少因数据划分不合理或偶然性因素带来的评估偏差。

### (1) 5 折交叉验证

使用五折交叉验证，首先对单一的机器学习模型进行交叉验证，然后在对使用 Stacking 算法集成的三种集成模型进行交叉验证，接着再对使用平均组合策略的集成模型进行交叉验证，使用 MAPE 作为模型评价指标。首先对单一模型进行五折交叉验证的实验结果如下表。

表 5.6 5 折交叉验证的单一模型 MAPE 均值

BP 神经网络	随机森林	支持向量回归	长短期记忆神经网络
2.96%	3.34%	4.19%	4.85%

从这个单一模型的五折交叉验证表可知，BP 神经网络模型对我国粮食产量预测的拟合效果最好。

然后是对使用堆叠算法与平均算法构建的模型进行交叉验证，五折交叉验证的实验结果如表 5.7 所示。

表 5.7 5 折交叉验证的 Stacking 算法与平均法集成模型 MAPE 均值

RF-BPNN-Stacking	RF-BPNN-SVR-Stacking	RF-BPNN-SVR-LSTM-Stacking	RF-BPNN-AVG	RF-BPNN-SVR-AVG	RF-BPNN-SVR-LSTM-AVG
2.67%	1.51%	1.57%	3.04%	3.74%	2.36%

表 5.7 列出的六种模型,这六种模型在 5 折交叉验证的结果显示使用 Stacking 集成后的随机森林-BPNN-SVR 模型的模型评价指标最好

### (2)10 折交叉验证

再进行了 5 折交叉验证之后,为了比较不同折数的交叉验证对模型的影响,我们接着使用 10 折交叉验证。首先,我们对四种机器学习模型进行 10 折交叉验证,交叉验证的结果如表表 5.8 所示。

表 5.8 10 折交叉验证的单一模型 MAPE 均值

BP 神经网络	随机森林	支持向量回归	长短期记忆神经网络
2.58%	2.64%	2.62%	2.74%

从表 5.8 的结果可以看出,对这四种机器学习模型进行 10 折交叉验证,模型的预测效果最好的是 BP 神经网络这与 5 折交叉验证的结果一致。

然后为了比较本文构建的 Stacking 集成模型,使用 10 折交叉验证对其进行验证,并且同时验证平均集成模型的效果。10 折交叉验证结果如表 5.9 所示。

表 5.9 10 折交叉验证的 Stacking 算法与平均法集成模型 MAPE 均值

RF-BPNN-Stacking	RF-BPNN-SVR-Stacking	RF-BPNN-SVR-LSTM-Stacking	RF-BPNN-AVG	RF-BPNN-SVR-AVG	RF-BPNN-SVR-LSTM-AVG
2.37%	1.71%	1.27%	2.61%	2.64%	2.17%

对这六种集成模型进行 10 折交叉验证,其中将 4 种机器学习模型用 Stacking 集成后的模型在其中的预测效果最好。

对本文构建的粮食产量预测模型进行了 5 折和 10 折交叉验证,验证结果表明使用 Stacking 集成算法的集成模型效果最好。而且在不同交叉验证方法下我们

的到的最优预测效果模型不一样。在 5 折交叉验证时 RF-BPNN-SVR-Stacking 预测效果最好，在 10 折交叉验证中，RF-BPNN-SVR-LSTM-Stacking 预测效果最好。

### 5.3 本章小结

本章主要是对本文构建的粮食产量预测模型进行分析，首先比较了四种机器学习模型与构建的集成模型的在不同预测年限的预测效果。然后使用不同数据集对构建的模型进行测试，说明模型在不同数据集上的预测效果较好，具有广泛性。接着再采用交叉验证的方式对本文构建的粮食产量预测模型进行验证，并分别在不同的交叉验证下验证。具体来说：BP 神经网络在预测短期的效果最好，而长短期记忆神经网络模型的预测的效果一般。SVR 模型在预测最近 3 年的数据时预测效果最好，但预测最近 1 年粮食产量数据时的效果一般。而 RFR 在 3 种不同时间段的预测效果都较好。在交叉验证的结果中，5 折交叉验证中，RF-BPNN-SVR-Stacking 拟合效果最优、10 折交叉验证中，RF-BPNN-SVR-LSTM-Stacking 拟合效果最优。

## 6 总结与展望

### 6.1 本文总结

粮食问题是党和国家高度重视的问题,解决吃饭问题一直在治理国家的首要位置。而粮食产量又受到国际国内因素影响,因此,科学预测粮食产量对于保障经济社会健康发展、粮食安全以及政府决策具有重要意义。本文使用机器学习方法对粮食产量进行建模分析,并且提出运用 Stacking 集成算法,对构建的 LSTM、BPNN、RFR 以及 SVR 模型进行融合。同时为了比较以往组合模型中常用的平均集成法,构建了平均集成模型用作对比模型。在构建粮食产量数据集前使用相关性分析法筛选出与粮食产量具有极强相关的影响因素。同时,选择相同变量的省域粮食产量数据集用作对比数据集,进行模型训练与评估。接着将构建了的 LSTM、BPNN、RFR 以及 SVR 等模型,进行粮食产量的预测并评估它们的预测效果。然后再用对比数据集测试模型在不同数据集上的预测效果,最后使用交叉分析验证各个模型的预测效果。对不同数据集上的预测结果表明,在使用省域数据集进行的模型预测效果与使用我国粮食产量数据与影响因素的数据集时存在差异,但总体上两种数据集都表现出较好的训练与测试效果。交叉验证的结果显示,Stacking 算法集成模型的预测效果优于平均法集成模型与单一模型的预测效果。在 5 折交叉验证中,RF-BPNN-SVR-Stacking 集成模型的预测效果最佳,而在 10 折交叉验证中,RF-BPNN-SVR-LSTM-Stacking 集成模型的预测效果最佳。

### 6.2 本文展望

使用 Stacking 算法构建的 3 种模型在对我国的粮食产量进行预测中取得了不错的预测效果,但是本文构建的集成预测模型还有一定的不足,第一在于初级学习器的选择上,每个初级学习器预测效果的优劣直接影响整体集成模型,因此,可以对选择的基学习器进行改进和优化,而且在初级学习器的选择上,也可用尝试使用不同的模型作为初级学习器进行模型组合;第二,在影响因素变量上,本文只采用了相关性分析进行变量筛选,未来可以用多一些变量筛选方法进行验证,

比如，使用主成分分析法或者主成分分析与相关性分析相结合来筛选特征变量，提高模型的预测效果；对模型的参数优化，本文构建的四种机器学习模型涉及多种参数调节，因此得出来的预测效果可能具有随机性，所以需要对这些参数进行优化，找到最合适的参数进行预测。

## 参考文献

- [1] Awad M M. Toward Precision in Crop Yield Estimation Using Remote Sensing and Optimization Techniques[J]. Agriculture,2019,9(3):54.
- [2] Doltra J ,Gallejones P ,Olesen J,et al.Simulating soil fertility management effects on crop yield and soil nitrogen dynamics in field trials under organic farming in Europe[J].Field Crops Research,2019,2331-11.
- [3] D.Maghesh Kumar,andDr.K.Mohan Kumar.decision support system for prediction of crop yield using re-engineered artificial neural network. Journal of Critical Reviews 2020,7(12):2290-2297.
- [4] Eatidal A,Luca P,Santiago B, et al.In-season forecasting of within-field grain yield from Sentinel-2 time series data[J].International Journal of Applied Earth Observation and Geoinformation,2024,126103636-.
- [5] García-León D ,Contreras S ,Hunink J .Comparison of meteorological and satellite-based drought indices as yield predictors of Spanish cereals[J].Agricultural Water Management,2019,213388-396.
- [6] Haider A S,Naqvi R S,Akram T, et al.LSTM Neural Network Based Forecasting Model for Wheat Production in Pakistan[J].Agronomy,2019,9(2):72.
- [7] Jiang S, Bai C F,Du C F.Traffic Forecast Based on Empirical Mode Decomposition and RBF Neural Network[J]. Advanced Materials Research, 2014,846-847:1270-1273.
- [8] Khaki S,Wang L, Archontoulis SV.A cnn-rnn framework for crop yield Frontiers in Plant Science,2020,10:1750.
- [9] Liu S,Baret F,Andrieu B,et al. Estimation of wheat plant density at early stages using high resolution imagery[J]. Frontiers in Plant Science, 2017,8739.
- [10] Omer Burstein, Tamir Grodek, Yehouda Enzel, et al. SatVITS-Flood: Satellite Vegetation Index Time Series Flood Detection Model for Hyperarid Regions[J]. Water Resources Research, 2023,9(2):72.

- [11] Pagani V, Guarneri T, Fumagalli D, et al. Improving cereal yield forecast in Europe The impact of weather extremes[J]. European Journal of Agronomy, 2017, 89:97-106.
- [12] Wang X, Huang J, Feng Q, et al. Winter Wheat Yield Prediction at County Level and Uncertainty Analysis in Main Wheat-Producing Regions of China with Deep Learning Approaches[J]. Remote Sensing, 2020, 12(11):1744.
- [13] 陈锡康. 全国粮食产量预测研究[J]. 中国科学院院刊, 1992, 12(04):330-333.
- [14] 陈全润, 杨翠红. 河南省粮食产量预测方法研究[J]. 系统科学, 2018, 38(07):8138-22.
- [15] 陈灿虎, 陈英伟. 基于差分进化改进灰狼优化的 BP 模型的全国粮食产量预测[J]. 农业与技术, 2023, 43(05):167-171.
- [16] 董美双, 何欢, 童晓星. 基于  $\epsilon$ -SVR 的粮食产量预测模型及应用[J]. 浙江大学学报(农业与生命科学版), 2009, 35(4):439-443.
- [17] 苟喻. 重庆市水稻估产要素及模型研究[D]. 西南大学学报, 2015(11).
- [18] 丁茜, 赵晓东, 吴鑫俊等. 基于 RBF 核的多分类 SVM 滑塌易发性评价模型[J]. 中国安全科学学报, 2022, 32(03):194-200.
- [19] 郭亚菲. 基于组合模型的粮食产量预测方法研究[D]. 河南工业大学, 2019(02).
- [20] 高心怡, 韩飞. 基于混合智能算法的支持向量机的粮食产量预测[J]. 江苏大学学报(自然科学版), 2020, 41(03):301-306.
- [21] 何延治. 基于时间序列分析的吉林省粮食产量预测模型[J]. 江苏农业科学, 2014, 42(10):478-479.
- [22] 胡程磊, 刘永华, 高菊玲. 基于 IPSO-BP 模型的粮食产量预测方法研究[J]. 中国农机化学报, 2021, 42(03):136-141.
- [23] 贾梦琦, 蔡振江, 胡建等. 基于机器学习的粮食产量预测模型研究[J]. 河北农业大学学报, 2021, 44(03):103-108.
- [24] 李国祥. 新中国解决粮食问题的 70 年探索[J]. 中国粮食经济, 2019(10):27-31.
- [25] 倪坤晓, 何安华. 中国粮食供需形势分析[J]. 世界农业, 2021(02):10-18.
- [26] 宋洪远. 实现粮食供求平衡? 保障国家粮食安全[J]. 南京农业大学学报(社会

- 科学版),2016,16(04):1-11+155.
- [27] 田稼科.基于时频分析和 LSTM 的粮食产量预测模型研究[D].河南工业大学.2021.
- [28] 王小航.基于长时间序列气象数据的小麦产量预测模型研究[D].山西农业大学学报,2021.
- [29] 吴越,张焕明.基于 Holt 两参数指数平滑法和 ARIMA 模型的长三角粮食产量的预测[J].武汉轻工大学学报,2020,39(01):30-36.
- [30] 邬粒,邹黎敏,周科.基于机器学习的重庆市粮食产量预测及影响因素分析[J].中国农机化学报,2023,44(10):185-193.
- [31] 邢聪仁.基于机器学习的安徽省粮食产量预测方法研究[D].安徽大学.2019(07).
- [32] 徐宁,李发东,张秋英等.基于机器学习和未来气候变化模式的埃塞俄比亚粮食产量预测[J/OL].中国生态农业学报(中英文),1-18[2024-01-04].
- [33] 薛腾,李保林,杨丽华.一种新颖的组合模型在粮食产量预测中的应用[J].科技创业月刊,2017,30(03):47-49.
- [34] 于润,洪欣,于泽翔等.基于 BP 神经网络的江苏省粮食产量预测[J].沈阳师范大学学报(自然科学版),2023,41(04):316-320.
- [35] 尤泰媛,程文仕.基于随机森林的粮食单产预测研究[J].国土与自然资源研究,2022,(04):50-54.
- [36] 张陆.我国粮食供需格局演变与开放条件下的粮食安全政策研究[J].农业经济,2023,(05):118-120.
- [37] 赵俊伟,王帅,王杰等.基于回归分析的粮食产量影响因素分析——以江西省为例[J].经济研究导刊,2013,(34):78-80+95.
- [38] 翟静,曹俊.基于时间序列 ARIMA 与 BP 神经网络的组合预测模型[J].统计与决策,2016(04):29-32.
- [39] 张岳,陈为真,陈梦娇.数据驱动的粮食产能组合预测模型[J/OL].南京信息工程大学学报(自然科学版),1-10[2024-01-13].
- [40] 庄星,韩飞.基于混合群智能算法优化 BP 神经网络的粮食产量预测[J].江苏

大学学报(自然科学版),2019,40(02):209-215.

[41] 张莹莹,徐文科.基于灰色 GM(1,1)模型对黑龙江省粮食产量的预测[J].哈尔滨师范大学自然科学学报,2019,35(03):41-45.

[42] 赵茹欣,王会肖,董宇轩.气候变化对关中地区粮食产量的影响及趋势分析[J].中国生态农业学报(中英文),2020,28(04):467-479.

## 致谢

三年时间仿佛转瞬即逝，转眼间，我已经度过了这段充实而宝贵的时光。在这短暂的三年里，我经历了无数的挑战和成长，见证了自己的变化和进步。回首往事，仿佛昨日，但内心却充满了对过去的感慨和对未来的期待。这三年让我学到了很多，收获了很多，也铭记了很多，它们将永远留在我生命的记忆里，成为我前行路上的宝贵财富。

此刻，我要向所有在我研究过程中给予帮助和支持的人们表示最诚挚的感谢。首先，我要向我的导师表达最深的感激和敬意。他不仅是我的学术指导者，更是我人生道路上的引路人和良师益友。导师的悉心指导、专业知识和无私支持，为我提供了坚实的学术基础和前行的动力。他的言传身教，教诲和启发，使我受益匪浅，受到了深刻的熏陶和影响。在我迷茫时给予我指引，在我困难时给予我支持，在我成功时给予我鼓励，导师的每一句话、每一个建议都铭刻在我的心中，我将永远铭记于心。

我还要感谢“咱的家”的师兄师姐师弟师妹们，感谢他们在我论文撰写中给予的指导与帮助，在数据处理过程中给予的帮助和指导。他们的经验和专业知识对我在研究中遇到的问题提供了宝贵的帮助，使我能够顺利地完成实证和数据分析。

此外，我要感谢我的家人和朋友们。感谢他们在我整个研究生阶段给予的理解、支持和鼓励。他们的陪伴和鼓励是我坚持不懈、克服困难的动力源泉。

最后，我要感谢所有在我研究过程中给予帮助和支持的老师、同学和其他相关人员。感谢他们在学术交流、实验设备使用等方面给予的帮助和指导。没有他们的支持，我无法顺利地完成这项研究工作。

在这篇论文的完成过程中，我收获了很多宝贵的经验和知识。这些经验和知识将对我未来的学习和工作产生重要的影响。我将继续努力学习，不断提高自己的专业能力，为社会做出更大的贡献。

再次向所有给予我帮助和支持的人们表示最衷心的感谢！

谨以此文献给所有关心和支持我的人们。

衷心感谢！