

项目编号:

暨南大学大学生创新创业训练计划 项目申报书

项目名称 基于我国粮食经济发展战略背景下
的小麦价格 ARIMA-LSTM-XGBoost 组合预测模型研究

项目类别 2022 年度大创计划项目申报

项目类型 创新训练项目

项目负责人 苏晓钰

学 号 2019052121

院（部）专业 经济学院经济统计学专业

入学年份 2019 年 9 月

联系电话 13690803668

E_mail 1670077624@qq.com

指导教师 林少萍

申报时间 2022 年 3 月

暨南大学教务处制

二〇二二年三月

填 写 须 知

一、项目类型分类说明：

1. 创新训练项目是本科生个人或团队，在导师指导下，自主完成创新性实验方法的设计、研究条件准备和项目实施、研究报告撰写、成果（学术）交流等工作。

2. 创业训练项目是本科生团队，在导师指导下，团队中每个学生在项目实施过程中承担一个或多个具体的角色，通过编制商业计划书、开展可行性研究、模拟企业运行、参加企业实践、撰写创业报告等工作。

3. 创业实践项目是学生团队，在学校导师和企业导师共同指导下，采用前期创新训练项目（或创新性实验）的成果，提出一项具有市场前景的创新性产品或者服务，以此为基础开展创业实践活动。

二、申报书请按顺序逐项填写，填写内容必须实事求是，表达明确严谨。空缺项要填“无”。

三、填写时可以改变字体大小等，但要确保表格的样式没有被改变；填写完后用 A4 纸张双面打印，不得随意涂改。

四、申报过程有不明事宜，请与教务处实践教学科联系和咨询。

项目名称		基于我国粮食经济发展战略背景下的小麦价格ARIMA-LSTM-XGBoost 组合预测模型研究		项目类型	一般项目	项目研究周期	一年
负责人情况	姓 名	苏晓钰	性 别	女	平均绩点	3.72	
	学 院	经济学院	专 业	经济统计学	入学年份	2019.09	
	联系电话	13690803668	QQ 号	1670077624	Email	1670077624@qq.com	
项目组成员 (不含负责人)	姓 名	性别	学号	院 (部)	专 业	入学年份	
	陈嘉好	女	2019051643	国际学院	金融学	2019.09	
	李敏怡	女	2019052020	经济学院	经济统计学	2019.09	
指导老师情况 (第一导师)	姓名	林少萍	学科专业	应用统计学	所在单位	经济学院统计学系	
	职 称 / 职务	讲师	学历学位	博士研究生	联系电话	13802906815	
校外导师情况 (创业实践项目必填)	姓名	无	所在单位		无		
	职 称 / 职务	无	学历学位	无	联系电话	无	

一、项目简介

国家粮食价格既是调节粮食产品供求关系的信号和手段，也关系到老百姓生活成本和农民收入水平。中国的小麦作为三大主粮之一，年产量超 1 亿吨，占据世界市场近 25%，处于世界粮仓的重要战略地位。然而自 2018 年起，中美贸易摩擦不断，为中国小麦进口带来更多不确定性；2020 年新冠疫情袭来，国内农产品生产受阻，小麦价格一度上涨；而国际疫情反复不断引起的市场波动、俄乌冲突动摇了乌克兰作为“欧洲粮仓”的重要地位，推动国际小麦价格，加剧了国际市场价格波动的风险。而国内研究的研究范围和研究方法存在局限性，大多仅使用传统的时间序列分析进行预测，且很少将国际动荡事件纳入影响因素，结论过时。在当前国家粮食经济发展战略背景下，构建更精准的小麦期货价格预测模型有利于更有效地反应小麦市场，保障小麦最低收购价政策顺利实施。

本项目基于《中国农产品价格调查年鉴》、《中国统计年鉴》与国家粮油信息中心的公开数据，从小麦国际和国内市场角度出发，探求影响小麦价格的重要因素，旨在对小麦发展趋势做出准确的预测。首先根据训练集数据，基于随机森林特征重要性排序构建了全面的小麦期货价格的影响因素指标体系，并构建出传统的时间序列分析模型与单变量的 LSTM 模型与单变量 XGBoost 模型。为了进一步构建更准确的小麦期货价格预测模型，我们基于所建立的影响因素指标体系，构建多变量的 LSTM 模型与多变量 XGBoost 模型。最后，基于 Stacking 集成学习算法改进现有的预测模型，考虑构建 ARIMA、多变量的 XGBoost 模型和多变量的 LSTM 模型的组合模型：将三组模型的预测值作为支持向量机回归 SVR 模型的输入，并采用贝叶斯优化算法对 SVR 模型进行调参，解决影响因素的未来值的获取问题并显著提高预测精度。综合组合模型与单一模型、单变量模型与多变量模型的预测结果，从绝对误差、相对误差、平均相对误差等指标比较模型，并选择出预测效果最好的模型。基于该模型的有效预测结果，提出相应政策建议，有助于及时调节市场供求关系，稳定粮食经济；将该模型进一步推广到农产品价格预测领域，对我国农业经济发展有重大意义。

二、立项依据（研究意义、国内外研究现状及发展动态分析等）

2.1 研究背景

2.1.1 国家战略

粮食是人类发展根基，粮食安全关乎国际民生，无论是粮食生产、储备还是粮食贸易历来备受各国重视，是国家发展的重中之重。中国，作为发展中国家中的人口大国，耕地资源低于世界 10%，是粮食生产大国，更是消费大国。^[1] 粮食问题是关系全局的重大战略问题，解决粮食问题、发展社会经济，粮食价格是最主要的关注点之一。党的十九大报告中强调：要确保国家粮食安全，把中国人的饭碗牢牢端在自己手中。中央一号文件连续十二年关注“三农问题”，并提出“发展农产品期货交易”。2016 年，明确提出了探索建立农业补贴、涉农信贷、农产品期货与农业保险联动相关机制，发挥粮食期货稳定产销关系、控制价格变动风险的作用。^[2]

小麦，作为三大主粮之一，全球年均约 75%用于口粮消费，其余作饲料、工业等用途，而中国的小麦年产量和消费量都居世界前列。2006 年起，我国对小麦实行最低收购价政策，稳定主粮价格、改善供求关系以保障农民利益，从而提高农民生产积极性。2017 年中央一号文件提出“要坚持并完善稻谷、小麦最低收购价政策，合理调整最低收购水平，形成合理比价关系”。^[3] 在近十年的国家政策扶持下，国产小麦产量提升、粮食库存充足，有效保障市场需求，小麦现货价格稳步上涨。

2.1.2 国内市场

中国小麦年总产量超 1 亿吨，年播种面积和产量均约占世界粮食常年种植面积和产量的四分之一，自给率达到 95%以上，在国民经济中占比巨大。^[5] 2020 年小麦种植面积 23380 千公顷，2021 年小麦总产量达 13694.6 万吨，同比增加 269.6 万吨，增长率 2%。同时，随着人口增长和经济发展，小麦需求量逐年上涨，市场对于优质高等小麦和饲料级小麦的需求旺盛。

纵观 2007 年世界粮食危机以来，得益于临储和“托市”政策，我国小麦价格从 2009 年 1 月的 1790 元/吨上涨至 2013 年 9 月的 2519.29 元/吨，上涨幅度约 40.74%。2013 年以后，小麦价格波动明显且呈现出跌涨互现趋势。^[5] 2020 至今，受外界多重因素影响，粮食稳定成为国家安全战略部署重要手段，国内小麦减产、储备性需求激增导致价格明显上涨，到 2021 年 12 月 15 日小麦月度价格超过往年同期，达到了 2856.67 元/吨；今年第一季度，小麦每吨价格突破 3000 元大关。

而国内中筋小麦品种占比超 55%，是国产小麦重要来源。由于优质高筋、低小麦被排除在最低价补贴政策之外，普麦与优质麦价格差距小，农民普遍选择种植性价比更高的普麦，进口成为补充我国优质小麦缺口的重要途径。^[6] 此前，由于国内小麦生产成本明显高于国际水平，小麦价格一直呈现国内高于国外到岸价的“倒挂”现象，2020 年美国墨西哥湾硬红冬麦配额内到岸税后价约每吨 1965 元，低于国内 220 元。^[7]

2.1.3 国际市场

小麦是国际农产品贸易中最为活跃、用途最广的农产品之一，其种植分布、消费及总贸易额占世界粮食作物第一，作为大宗粮食作物占世界贸易近 46%。^[6] 国际上呈现出口国集中，进口国分散的局面，总体来看各地区进口来源国较为固定，其中美国是小麦第一大出口国。

随着人口增长以及人们对于优质小麦日益增长的消费需求，自 2001 年起，我国逐步从小麦净出口国转变为净进口国，主要进口优质小麦和饲料用小麦以调剂粮食结构。2021 年小麦进口

量 977 万吨创新高，虽仅占我国总量约 3%，但仍为世界小麦进口大国。长期以来，美国、加拿大和澳大利亚进口小麦占我国进口量八成以上，依存度高，存在一定的粮食安全隐忧；2018 年起，中美贸易摩擦不断，为中国小麦进口带来更多不确定性。而 2007 年和 2011 年两次粮食危机中，乌克兰和俄罗斯两个小麦出口大国采取了严厉的出口限制措施，进一步推高国际粮食价格，造成国际市场动荡。^[8]

随着一带一路倡议提出，沿线国家经济领域合作加深，由于俄罗斯和乌克兰小麦产量占全球 14%，出口总额近 30%，中国和俄罗斯、乌克兰、哈萨克斯坦的小麦贸易关系逐步建立，寻求多元的进口途径，降低对少部分国家的粮食依存度。^[9] 本月，在国际小麦价格暴涨的形势下，以 3 月 3 日为例，国内小麦现货价达 3210.67 元/吨，进口美国小麦完税成本突破 3600 元/吨，价差倒挂每吨 400 元左右，进口美国小麦已无成本优势。期间，国内发布公告“允许俄罗斯全境小麦进口”，改善粮食储备结构。

2.1.4 外部影响

(1) 2020 新冠肺炎疫情袭来，WHO 将其定义为全球大流行病，当前仍在全球范围蔓延。最新《全球粮食危机报告》指出，新冠肺炎疫情大流行或导致全球遭受严重饥饿人口数量翻 1 倍，粮食安全成为关注焦点。^[10] 疫情爆发初期，国内受疫情影响，交通封锁、停工停产等使粮食运输受限导致原料短缺、面粉生产加工停摆，加之人们在封锁隔离期恐慌性储备主粮，国内小麦价格短期内一度上涨。随着疫情转移到国外，2020 年 7 月以后，国际谷物价格进入了上升通道。同年 10 月 20 日，芝加哥期货交易所（CBOT）小麦主力合约报收价比月初上涨了 11.2%。^[7] 受疫情影响，各国封锁下进出口贸易受阻，外部供给环境恶化；谷物出口大国担忧粮食供应，相继限制粮食出口；同时发达国家采取量化宽松政策刺激经济推高大宗商品价格；在经济衰退下，投资者纷纷将目光转向粮食市场，多重作用下，国际小麦价格上涨。由于国内小麦现价与国际价格紧密关联，国际疫情引起的市场波动往往对我国小麦生产销售市场产生影响。

(2) 粮食作物产量和质量直接影响现期货价格。同一时期，非洲蝗灾、美欧小麦主产区遇寒潮，影响小麦优良率和作物生长状况，降低新麦上市预期。^[7] 国内，华北地区秋收秋种期遭遇持续降雨，部分地区小麦受损质量下降、播种推迟增加成本，引发市场对小麦减产担忧，支撑小麦价格高位运行。^[11]

(3) 乌克兰作为“欧洲粮仓”，与俄罗斯小麦产量皆占世界前五。随着俄乌冲突升级，欧美对俄罗斯制裁，加剧疫情后复苏带来的通胀压力。一方面，国际市场尤其是依赖两国出口小麦的国家对于未来粮食供应状况充满忧虑，且黑海港口贸易陷入混乱，增大粮食运输阻力；另一方面，小麦期货作为金融产品，除了作为现货价格风向标，在特殊时期全球资本涌入市场避险、投机。半月内，伴随国际油价上涨，推动小麦价格更上一层楼，CBOT 小麦期货价格涨幅 40%，为 14 年最高。^[12] 2 月 21~25 日，中国小麦进口成本水涨船高，普遍上涨 10%以上。^[13] 当下，中国开放俄罗斯小麦进口，节省海运成本，以较低价格保障国家小麦库存量，同时也借此机会加快进口小麦结构调整，降低对特定国家依存度。

总体而言，商品的供求及价格决定，即市场中的均衡分析是经济分析中最基本的问题之一，小麦作为中国乃至世界主粮，其现货价格变动以及市场动态备受各方关注。

2.2 研究理论价值与实践意义

2.2.1 理论价值

本课题基于 stacking 集成学习算法, 构建 ARIMA、多变量的 XGBoost 模型和多变量的 LSTM 模型的组合模型, 克服了单变量预测模型的片面和预测准确率偏低的情况, 同时对单一模型的预测结果进行了优化, 具有一定的创新意义, 丰富了农产品价格预测模型。

在数据的实证分析方面, 本文采用了 2009 年 7 月 1 日至 2022 年 3 月 3 日的日数据, 其中包含了 18 年中美贸易战、19 年年末至今的全球新冠肺炎流行病、22 年 1 月至今的俄乌冲突等有可能使小麦价格产生大幅波动的关键时间节点, 弥补了在国际冲突不断的后疫情时代下关于小麦期货价格预测实证研究的空白。

2.2.2 实践意义

(1) 调节市场供求关系

价格受供求关系影响, 又反作用于供求关系。过去 20 年, 我国小麦产量稳步提升, 基本满足国民消费需求, 粮食库存充足, 同时受政策影响价格涨幅较大。由于国内小麦生产从成本高, 国家通过小麦最低收购价补贴、临储麦政策“托市”, 维持了小麦供应量, 从而保障消费需求。

在此期间, 我国粮食期货市场日臻成熟, 小麦期货价格成为市场价格“风向标”。作为大宗商品, 小麦期货为农民和供应链下游的加工商、贸易商提供套期保值的功能。通过期货市场和现货市场的信息传递, 有效反应市场当前和预期供求关系。合理的价格预测, 一方面, 指导农民调整小麦种植结构、合理安排售粮时间, 保障粮食供给、控制供求不平衡带来的损失; 另一方面, 对于供应链下游, 有助于生产商、销售商根据预期价格趋势进行采购、生产和存货管理。^[2] 通过有效价格预测, 减少卖方与市场的信息不对称, 提高整条小麦供应链效率、保障生产者利益, 满足市场需求, 构建健康的主粮市场。

(2) 规范国内小麦定价

2006 年实行小麦最低收购价以来, 国产小麦价格增长近一倍。由于长期以来粗放型的生产方式导致我国小麦种植成本高于国际水平, 为保障国家粮食安全, 国家政策补贴在一定程度上扭曲小麦价格。2016 年对于粮价补贴政策的改革在一定程度上提升了市场对于小麦价格的调节能力, 此后我国小麦价格一直在小范围波动上涨, 更加有效地反应市场供求关系。

此外, 作为粮食进口大国, 我国小麦贸易逆差大、买方市场势力相对较弱, 对于粮食定价影响力低。^[14] 相比于出口大国对于小麦种植的扶持力度, 我国还有一定距离。尽管国内外价差有缩小趋势, 但今年三月以前, 进口小麦价格一直优于国内, 国产小麦价格及质量竞争力较低。对小麦价格的有效预测, 是国家加快推进小麦和稻谷定价机制、进行补贴政策 and 收储体制改革的重要依据; 与此同时, 帮助人们有效应对国际粮价波动带来的风险, 促进国产小麦价格与国际化、标准化接轨, 提升我国小麦产业国际影响力。

(3) 保障国家粮食安全

后疫情时代, 在小麦价格波动剧烈、市场风险加剧的整体经济大环境下, 准确预测小麦现货价格有利于降低外部环境对于粮食供应、粮食价格的冲击, 优化粮食储备。尤其是当下疫情和政治等不确定性因素加剧, 小麦现货市场动态是外界环境的“晴雨表”。

作为金融期货市场的物质基础, 小麦期货市场围绕现货价格波动。近期小麦期货价格的暴涨, 在一定程度上反映了主粮, 这类大宗商品, 其避险属性突出; 此外, 众多投机者借助国际

事件涌入小麦期货市场以期进行无风险套利，同时借机推高国际利率，收割高额回报。^[15]这一特殊时期，小麦价格的合理预测有助于稳定现货价格，缓解全球性恐慌造成的价格虚高，平衡期现货市场从而避免投机行为的出现。小麦作为国家主粮，其作用应该是满足人们口粮需求、确保社会生产有序，而非投机者哄抬物价获利的工具。因此，掌握未来价格趋势的对于小麦发挥粮食作用、国家粮食供应稳定具有重要意义。

2.3 国内外研究现状

2.3.1 国内外研究现状

小麦价格受到供给、需求、政策导向、替代品价格等因素的影响，导致价格波动频繁。下列为国内外学者关于小麦价格的预测方法的探究。

国外学者的研究主要集中在小麦价格的预测方法上：Haile(2016)^[16]等通过采用新开发的多国作物日历特定季节分解模型估算了小麦全球总供应量，并相应地预测出其价格变化和价格波动。研究表明，产出价格波动与小麦供应呈负相关，价格风险降低了小麦的生产反应，从而抑制了价格激励效应。Pal(2018)^[17]等通过对原油价格和粮食价格进行交叉相关分析，证明了原油价格与世界粮食价格之间存在强烈的正向依存关系；Ahumada(2016)^[18]等利用向量自回归方法(VAR)建立小麦、玉米和大豆三者价格交叉影响的价格预测模型，结果表明三者互为替代品，价格存在此消彼长的关系。

国内关于小麦价格预测主要集中在影响因素上：如柳燕子(2018)^[19]基于2009至2017年全国小麦收购价格指数，针对价格序列季节性、周期性、非线性及多尺度性等特征，提出基于EEMD分解、ARIMA预测的多模型小麦价格预测方法，为小麦价格预测提供了新思路。李雪(2018)^[20]等将最低收购价政策作为虚拟变量引入GARCH模型，实证研究政策对平抑小麦期现货价格波动的作用，结果表明，政策对小麦期现货价格波动产生了明显的影响，能够显著降低小麦现货市场的波动程度，但政策对期货价格波动的作用方向则相反。张瑞娟和高芸(2016)^[21]通过VAR模型和格兰杰因果检验分析了小麦最低收购价格和国际小麦月度价格对我国小麦价格的影响，提出国际小麦价格与我国小麦价格变化没有因果关系，而小麦最低收购价是拉高国内小麦价格的主要因素。崔海莹(2022)^[22]等通过构建DCC-MGARCH模型对多个时间序列之间的动态相关性进行估计，分析新冠肺炎疫情冲击对于三大主粮价格的影响效应和作用机制，认为的国内粮价会受到国际传导作用，并通过“国际疫情恶化-国际粮食价格波动-国内粮食价格波动”的传导机制产生影响。

2.4 文献述评

总体而言，国内外学者关于小麦价格的研究主要从探寻价格影响因素、研究价格波动走势、开展小麦价格与其他重要物资价格影响机制等角度的相关研究。主要运用的模型有向量自回归方法(VAR)、定季节分解模型、BP神经网络(BPNN)算法、ARIMA预测等进行定量研究，尤以VAR模型使用最广，但前三种方法所需要的指标较多，且对于我国目前的统计发展现状而言，大部分指标仍难以大范围获取。

同时，现有文献主要采用定性的方法，研究新冠肺炎疫情对中国粮食价格的影响、习惯将小麦价格与其它粮食关联，但对中国粮食价格的影响效应及作用机制没有得到充分关注，从而忽略了小麦自身价格属性和变动情况。同时，在后疫情时代，粮食价格波动情况不同于以往，

对于这一时期小麦现货价格的研究还处于未成熟阶段。笔者认为对于粮食价格预测的模型可能受外部因素影响而使准确率发生变化，采用疫情前数据进行研究的模型需适时改进。

表 3-1 不同预测模型间的优缺点比较

模型	优点	缺点
ARMA	模型简单方便	只可对平稳序列进行预测
AMIMA	模型简单，只需借助内生变量	只能捕捉线性关系
SVR	可解决非线性、高维问题	数据量大时训练速度较慢
神经网络（LSTM）	准确度高，能充分逼近复杂的非线性关系	需要大量的参数，学习时间过长
灰色预测（GM）	所需要的样本数据量小，且对复杂系统的预测效果较好	只适合指数增长和中短期的预测

2.5 参考文献

- [1]朱晶,张庆萍.中国利用俄罗斯、乌克兰和哈萨克斯坦小麦市场分析[J].农业经济问题,2014,35(04):42-50+111.DOI:10.13246/j.cnki.iae.2014.04.007.
- [2]杨惠珍,韦敬楠,张立中.我国粮食期货市场价格发现功能的实证分析——以玉米和小麦市场为例[J].价格月刊,2017(05):19-23.DOI:10.14076/j.issn.1006-2025.2017.05.04.
- [3]刘平,方旖旎.中国小麦进口结构及竞争力比较研究[J].市场周刊,2021,34(08):144-146.
- [4]高云,孙一铮,郭新宇,矫健.国内外小麦价格相关性及其影响因素研究[J].价格理论与实践,2018(05):71-74.DOI:10.19851/j.cnki.cn11-1010/f.2018.05.018.
- [5]林春霞.国际粮价将呈高位震荡趋势[N].中国经济时报,2021-09-29(001).DOI:10.28427/n.cnki.njjsb.2021.002172.
- [6]赵荣,曹洁,朱婷婷.浅议中国小麦的国际贸易及竞争力提升[J].滁州学院学报,2021,23(01):14-19+41.
- [7]朱聪,曲春红,王永春,赵伟.新一轮国际粮食价格上涨:原因及对中国市场的影响[J/OL].中国农业资源与区划:1-13[2022-03-08].<http://kns.cnki.net/kcms/detail/11.3513.S.20211130.1225.010.html>.
- [8]钟钰,陈希,牛坤玉.粮食出口限制政策的实施效果与我国应对——来自部分小麦出口国的证据[J].经济纵横,2021(08):29-39+137.DOI:10.16528/j.cnki.22-1054/f.202108029.
- [9]潘寅茹.乌克兰为何被称为“欧洲粮仓”?玉米小麦出口均占全球10%以上[N].第一财经日报,2022-03-04(010).
- [10]曹慧.新冠肺炎疫情对国内外小麦产业的冲击:趋势、问题与建议[J].世界农业,2021(01):4-10+27.DOI:10.13856/j.cnki-1097/s.2021.01.001.
- [11]申洪源.2021年中国小麦市场分析[N].粮油市场报,2022-02-15(B02).DOI:10.28553/n.cnki.nlysc.2022.000245.

- [12]邵海鹏. 小麦也“疯狂”: 进口首次突破配额, 价格站上历史高位[N]. 第一财经日报, 2022-03-02(A06).
- [13]邵海鹏. CBOT 小麦期货连续涨停, 国内外小麦价格倒挂影响几何[N]. 第一财经日报, 2022-03-04(A10).
- [14]孙致陆. 贸易开放背景下国际小麦贸易市场势力实证分析[J]. 华中农业大学学报(社会科学版), 2019(04):1-14+169. DOI:10.13300/j.cnki.hnwkxb.2019.04.001.
- [15]陈植. 芝商所小麦玉米期货迭创新高 对冲基金押宝“新避险投资”[N]. 21 世纪经济报道, 2022-03-04(005).
- [16]Haile M G, Kalkuhl M. Braun J V Worldwide Acreage and Yield Response to International Price Change and Volatility: A Dynamic Panel Data Analysis for Wheat, Rice, Corn, and Soybeans M/Food Price Volatility and Its Implications for Food Security and Policy. Springer International Publishing, 2016
- [17]Pal D, Mitra S K Interdependence between crude oil and world food prices: A detrended cross correlation analysis. Physica A Statistical Mechanics & Its Applications, 2018
- [18]Ahumada H, Cornejo M Forecasting food prices: The case of corn, soybeans and wheatly] International Journal of Forecasting. 2016.
- [19]柳燕子. 基于多模型分析的小麦价格预测研究[J]. 山西农经, 2018(04):148. DOI:10.16675/j.cnki.cn14-1065/f.2018.04.103.
- [20]李雪, 韩一军, 付文阁. 最低收购价政策对小麦市场价格波动影响的实证分析[J]. 华中农业大学学报(社会科学版), 2018(02):1-7+154. DOI:10.13300/j.cnki.hnwkxb.2018.02.001.
- [21]张瑞娟, 高芸. 国内外小麦价格联动关系研究[J]. 价格理论与实践, 2016(07):112-115. DOI:10.19851/j.cnki.cn11-1010/f.2016.07.029.
- [22]新冠肺炎疫情对中国粮食价格的影响效应及作用机制. 崔海莹

三、项目方案（具体研究内容、研究目标、计划、技术路线、人员分工等）

3.1 具体研究内容

本项目基于《中国农产品价格调查年鉴》中的 2009 年 1 月 4 日至 2021 年 12 月 17 日共 3255 个样本我国小麦期货价格序列的日数据、以及《中国统计年鉴》、国家粮油信息中心关于农业经济投入、人力投入、我国财政收支、小麦对外贸易、小麦国际进出口量等多个影响因素, 建立数学模型预测小麦期货价格。

第一部分：构成小麦期货价格的影响因素指标体系

计算 Pearson 相关系数度量变量之间的非线性相关程度通过随机森林特征重要性排序, 最终筛选出少数具有影响力的因素, 构成预测小麦期货价格的指标体系,

第二部分：基于所构建的小麦期货价格的预测指标体系, 建立模型进行预测

(1) 通过传统的 ARIMA 模型建立经典的单变量时间序列模型;

(2) 基于 XGBoost 模型和 LSTM 模型将无监督的时间序列数据进行转化, 建立预测市场现货价格的单变量 XGBoost 模型和单变量 LSTM 模型;

(3) 为了提升模型预测效果, 根据由随机森林提取的特征构建影响因素指标体系, 分别构

建多变量的 XGBoost 模型和多变量的 LSTM 模型并进行预测；

(4) 参考 Stacking 集成学习算法在多领域中取得的丰硕成果，考虑构建 ARIMA、多变量的 XGBoost 模型和多变量的 LSTM 模型的组合模型。第一步，基于 ARIMA 模型对各个影响因素的未来值进行预测；接着，将指标体系中的预测值作为影响变量，构建出改进的多变量的 XGBoost 模型和多变量的 LSTM 模型，最终将传统 ARIMA 模型以及上述改进模型所得的三组预测值输入支持向量机回归 SVR 模型，并采用贝叶斯优化算法调参，最终得到优化的组合模型的小麦期货价格预测值。

第三部分：评价单一预测模型与组合模型的预测效果，并进行最优选择

通过计算得到测试集的误差、相对误差以及平均相对误差，并比较 ARIMA-LSTM-XGBoost 组合模型与各单一预测模型如 ARIMA、单变量的 XGBoost 模型、LSTM 模型等预测模型下的指标，最终选出最适合的预测模型，并应用于我国小麦市场发展趋势预测。

3.2 研究目标

(1) 了解国内小麦市场的粮食农业扶持政策、市场发展现状和了解小麦市场风险，深入挖掘国内小麦价格的重要影响因素。

(2) 基于近十二年各个维度的数据，构建多个小麦期货价格预测模型。

(3) 比较 ARIMA 模型、单变量 XGBoost 模型和单变量 LSTM 模型、多变量 XGBoost 模型和多变量 LSTM 模型、ARIMA-LSTM-XGBoost 的预测效果，评价并选择出效果最好的预测模型

(4) 将本课题的预测模型应用到农业经济领域中，并利用该模型对未来农业布局、农产品供需关系的探讨和相关政策提出有价值的建议

3.3 研究计划

第一步：确定课题研究方向。了解当前国内小麦市场影响因收集宏观粮食农业扶持改革政策、国内外政治、经济、社会影响因子的资料，了解小麦市场风险等情况、市场概况以及所面临的问题，引入了本可以的研究对象，通过团队讨论以及专家帮助确定研究问题和研究方向。

第二步：对经济统计领域的预测模型进行学习与研究。通过查阅文献了解当前农产品价格预测的研究现状，学习所使用的相关方法与原理，为后续分析小麦价格波动成因和价格预测奠定了理论基础。

第三步：基于专业的统计知识构建多个预测模型并检验预测效果，其中包括 ARIMA 模型、LSTM 模型、XGBoost 模型等。通过随机森林特征重要性排序挖掘出国内小麦价格的重要影响因素，并将所构建的指标体系运用到预测模型中，提高预测率。针对单一模型的缺陷进行改进，如，根据模型预测结果，对所使用的所有预测模型进行评价，并选出预测效果最好的模型。

第四步：整理结果，撰写报告。将本项目的成果以报告的方式呈现给中国农业农村部、中国粮油信息中心等相关部门，为国家农产品扶持政策的施行进言献策。

3.4 研究方法

3.4.1 ARIMA 模型

整合自回归滑动平均模型 (ARIMA, 即 Autoregressive Integrated Moving Average) 是自回归和滑动平均模型的综合，此模型由 Box 与 Jenkins 提出，通常表示为 $ARIMA(p, d, q)$ 和包括自回归 (AR) \ 整合 (I) 及移动平均 (MA) 分别记为自回归项 p ，移动平均项 q 和差分次数 d 。其中心思

想是运用数学模型近似描述一个随机的时间序列。ARIMA 模型是根据 ARMA 模型拓展衍生而来的，可以处理非平稳性的、具有线性趋势和季节、循环成分的时间序列。ARIMA 模型的结构为：

$$\begin{cases} \phi(B)\Delta^d x_t = \Theta(B)\varepsilon_t \\ E(\varepsilon_t) = 0 \\ Var(\varepsilon_t) = \sigma_s^2 \\ E(\varepsilon_t \varepsilon_s) = 0, s \neq t \\ E(x_t \varepsilon_s) = 0, \forall s < t \end{cases}$$

建立一个 ARIMA 模型需要解决以下 3 个问题：

- （1）将非平稳序列转化为平稳序列。
- （2）确定模型的形式。即模型属于 AR、MA、ARMA 中的哪一种。
- （3）确定变量的滞后阶数。

ARIMA 模型的建模预测主要包括以下步骤：

1. 通过时序图、ACF 和 PACF 图结合来判断序列是否平稳，同时通过 ADF 单位根检验法构造 ADF 统计量，进行平稳性检验。
2. 当序列不平稳时，为例消除显著的线性趋势和季节趋势的影响，我们对时间序列进行平稳化处理，如差分法、数据的分解与平滑、对数化等。
3. 通过自相关系数图 (Autocorrelation Function, ACF) 和偏自相关系数图 (Partial Autocorrelation Function, PACF) 定阶法，观察两图的截尾和拖尾情况，初步确定 p(P) 和 q(Q) 的阶数，并结合 AIC 信息准则找到适合模型的最佳阶数 P, Q 的值。AIC 准则是由 K-L 信息量推导而来，其一般形式为：

$$AIC = -2\ln(L) + 2m$$

L 是极大似然函数，m 是待估参数个数，AIC 值越小代表模型效果越好

表 3-2 ARIMA 模型定阶原则

模型定阶	AR(p)	MR(q)	ARMA(p,q)
自相关系数($\hat{\rho}(k)$)	拖尾	P 阶截尾	拖尾
偏自相关系数(ϕ_{kk})	Q 阶截尾	拖尾	拖尾

4. 考虑到小麦期货价格序列的季节效应、长期趋势效应和随机波动之间存在复杂的交互影响关系，简单的 ARIMA 模型不足以提取其中的相关关系，因此我们采用乘积季节模型对模型进行优化，根据最小 AIC 与似然比统计量确定模型。
5. 通过对残差序列的随机性检验与正态性检验评价模型是否拟合恰当，本项目将采取观察残差标准差图、残差序列 Q-Q 图、Ljung-Box 法来检验模型的显著性。Ljung-Box 法白噪声检验的假设条件为：

原假设 $H_0 : \rho(1) = \rho(2) = \dots = \rho(m) = 0, \forall m \geq 1$

备择假设 H_1 : 至少存在某个 $\rho(k) \neq 0, \forall m \geq 1, k \leq m$

检验统计量 Q_{LB} 取为：
$$Q_{LB} = n(n+2) \sum_{k=1}^m \left(\frac{\hat{\rho}^2(k)}{n-k} \right) \sim \chi^2(m), \forall m > 0$$

综上所述，本研究建立的 ARIMA 模型的应用步骤主要包括平稳性判别、模型定阶与识别、参数估计、模型的优化选择、模型的显著性检验、模型预测六个步骤。

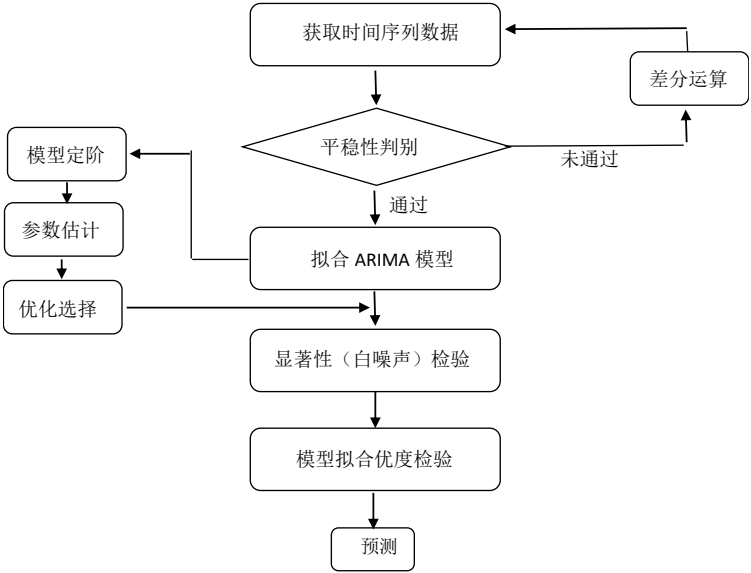


图 3-1 ARIMA 模型构建流程图

ARIMA 模型能较准确地预测小麦现货价格的时间序列整体趋势，但对价格时序局部细化的拟合效果不同，且部分拟合精度较低。该模型在外推时不方便加入外生变量，即没有考虑结构变动带来的影响，如政策因素、新冠疫情等突发情况等，而我国小麦现货价格是政策、市场等多方面因素综合影响下的结果，因此用 ARIMA 模型可能导致未来的预测效果不好。

3. 4. 2 LSTM 神经网络

本课题构建了单一隐形层，隐藏层节点数为 3 的 LSTM 模型。使用 MSE 作为损失函数，tanh 作为激活函数，以 Adam 算法作为优化器，branch size 为 1，迭代次数为 100 次，对训练集进行训练。长短期机器网络（Long Short-Term Memory, LSTM）是为了解决 RNN 存在梯度消失或梯度爆炸的问题而提出的特殊的 RNN 网络类型，基本结构与 RNN 相同。

RNN 结构如图所示。RNN 结构与传统神经网络相似，同样由输入层、隐含层和输出层组成，但神经元连接方式不同。RNN 的隐藏层神经元加入了权连接，使得隐含层神经元之间形成有序连接，每个神经元的输出都依赖于历史信息，实现时间序列动态信息在神经网络内部的转换，解决了传统神经网络无法对具有时间相关性的输入建模问题，从而更适合时间序列数据。

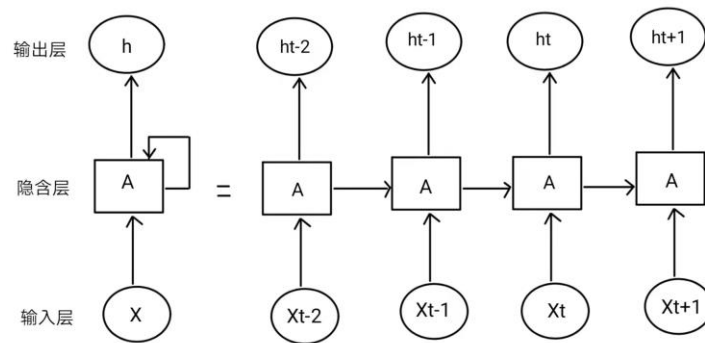


图 3-2 RNN 结构

但是在实际的序列数据的处理中，由于 RNN 在反向传播时采用按时间反向传播算法 (BPTT) 机制，可能会引发梯度消失或梯度爆炸的问题，而引入 LSTM 模型可以重新设置 RNN 的隐藏层模块结构，实现了对信息的长期记忆。LSTM 网络主要是通过引入门结构：遗忘门 (forget gates)、输入门 (input gates) 和输出门 (output gates)，对信息进行保护和控制。

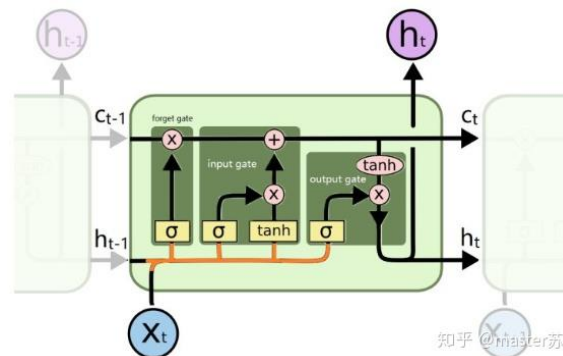


图 3-3 LSTM 网络单元结构

遗忘门是对于是否保留或丢弃上一时刻的 s_{t-1} 、 c_{t-1} 到当前时刻的状态 c_t 做出抉择。通过接收上一个状态的输出值 h_{t-1} 和当前时刻的输入 x_t ，使用 sigmoid 激活函数得到 0-1 之间的值，其中 1 表示全保留，0 表示全忘记，具体计算公式为：

$$F_t = \sigma(X_t W_{ih}^F + H_{t-1} W_{hh}^F + b_{ih}^F)$$

其中 W 代表连接两层的

输入门的作用是判断 c_t 要留下多少当前的输入信息 x_t ，共同更新当前时刻的单元状态。首先第一层由 sigmoid 层判断需要更新的输入，即

$$I_t = \sigma(X_t W_{ih}^I + H_{t-1} W_{hh}^I + b_{ih}^I)$$

第二层由 tanh 层创建可以更新到当前状态的候选信息，即

$$\tilde{C}_t = \tanh(X_t W_{ih}^{\tilde{C}} + H_{t-1} W_{hh}^{\tilde{C}} + b_{ih}^{\tilde{C}})$$

根据遗忘门和输出门两个门结构的计算可以得到记忆细胞 c_t 的更新值：

$$C_t = F \square C_{t-1} + I_t \square \tilde{C}_t$$

输出门的功能是决定当前时刻 c_t 中可以输出到 s_t 的部分，由通过 sigmoid 层得到的中间变量 o_t 与经过 tanh 层处理后得到 C_t 决定：

$$O_t = \sigma(X_t W_{ih}^O + H_{t-1} W_{hh}^O + b_{ih}^O)$$

$$s_t = o_t \circ \tanh(C_t)$$

对得到的训练集预测值进行反归一化可以得到 2009.07-2021.07 年真正的预测数据，并绘制相应的拟合曲线。

3.4.3 XGBoost 模型

极端梯度提升 (Extreme Gradient Boosting, XGBoost) 模型是一种分布式高效梯度提升算法，它是一种有监督学习算法，可解决分类、回归等机器学习问题，按照基学习器分类可以分为树模型和线性模型两种。XGBoost 树模型是由多棵 CART 决策树作为子模型构成的集成模型，采用 Gradient Tree Boosting 算法对多棵 CART 决策树进行集成学习，每棵决策树预测真实值和之前全部决策树预测值之和的残差，最终结果为全部决策树的预测值累加之和。

3.4.4 Stacking 集成学习

集成学习 (Ensemble learning) 的本质就是针对同一问题，通过采用某种策略整合数个有差异的个体学习器学习得到的结果，从而获得一个优于单个个体学习器预测性能的集成模型。集成学习模型可以分为同质集成和异质集成。典型的同质集成模型包括 Bagging 和 Boosting 等，Stacking 集成学习方法则属于异质集成模型。

随着近年来集成学习算法的研究愈发深入，Stacking 集成学习方法开始被广泛使用。与 Bagging 和 Boosting 算法不同的是，Stacking 是对不同的学习模型进行组合，并且采用了元学习模型代替了投票法。元学习模型的输入即个体学习器，也就是层模型的输出，元学习模型的输入特征数即所使用的层模型的个数。Stacking 算法其实就是通过建立多个层模型得到不同模型的预测结果，然后将其作为输入信息输入到下一层的学习器（元学习模型）中，使得该层的学习器能够对上一层学习器的成果充分学习，进而得到更高的预测精度，其流程图如下图所示。需要注意的是，Stacking 在对个体学习器进行选择时，应尽量选择预测效果好且异质性较强的模型，这样通过元学习模型的集成学习才容易取得更好的预测成果。

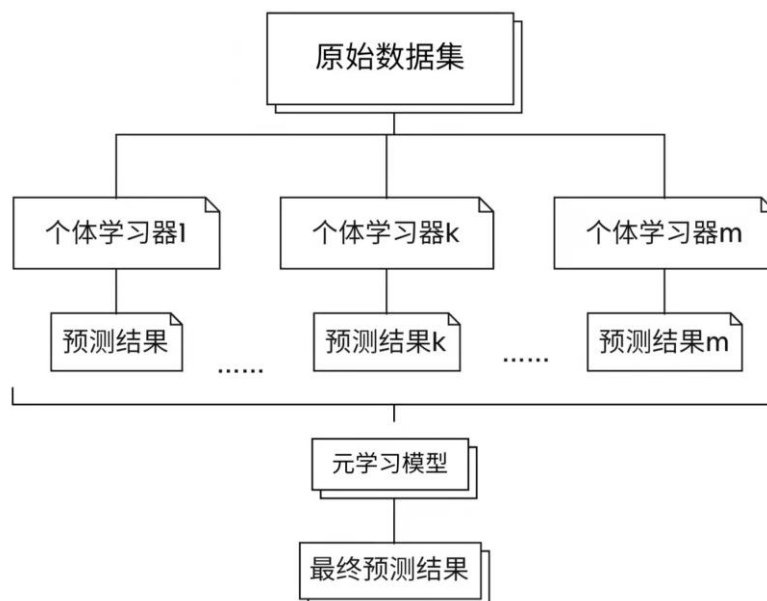


图 3-5 集成学习流程图

3.4.5 贝叶斯优化算法

在模型拟合时，选择合适的参数以实现高精度并非总是容易的，既可能出现拟合过度，也可能出现拟合不足，不同的参数设置可能会导致模型性能上的显著差异。贝叶斯优化算法正是为了寻找最优的参数组合而提出的，近年来，它在机器学习超参数调优方面得到了广泛的应用。

与网格搜索和随机搜索优化算法相比，贝叶斯优化算法能找到最优的超参数集，具有更高的搜索效率，对非凸问题进行调参的结果仍然是稳健的；同时，它能在开采和勘探之间找到一个平衡点来避免算法陷入局部最优，使得模型的性能达到全局最优²¹。

基于 python 软件的 hyperopt 模块实现贝叶斯优化算法主要包含以下四个部分：

- (1) 定义目标函数。目标函数即为我们想要使其达到最小化的对象，一般使用当前参数下的数据集的损失函数。
- (2) 设定参数搜索空间。在目标函数确定后，应对参数的搜索范围和分布进行设定，使算法在该区间下寻找是定义的目标函数达到最小的参数最优组合。
- (3) 优化算法的选择。优化算法的作用为在构造替代函数后选择下一组超参数组合，基于目标函数进行评估，主要包括：随机搜索，模拟退火以及 TPE 算法等。
- (4) 结果的历史数据。在此部分中，能够记录并存储历史选择的超参数组合和所定义的目标函数值，也就是我们想要最小化的函数值。

3.4.6 模型比较

通过计算得到测试集的误差、相对误差以及平均相对误差，并比较不同的预测模型下的几个指标，可以选择出预测精度较高的小麦期货价格预测模型。

表 3-3 单一模型与组合模型的预测结果比较

模型	年份	真实值	预测值	误差	相对误差	平均相对误差
ARIMA 模型	2021. 08					
	a%
	2022. 07					
单变量 LSTM 模型	2021. 08					
	b%
	2022. 07					
单变量 XGBoost 模型	2021. 08					
	c%
	2022. 07					
多变量 LSTM 模型	2021. 08					
	d%
	2022. 07					
多变量 XGBoost 模型	2021. 08					
	e%
	2022. 07					
ARIMA-LSTM-XGBoost 组合模型	2021. 08					
	f%
	2022. 07					

3.5 技术路线图

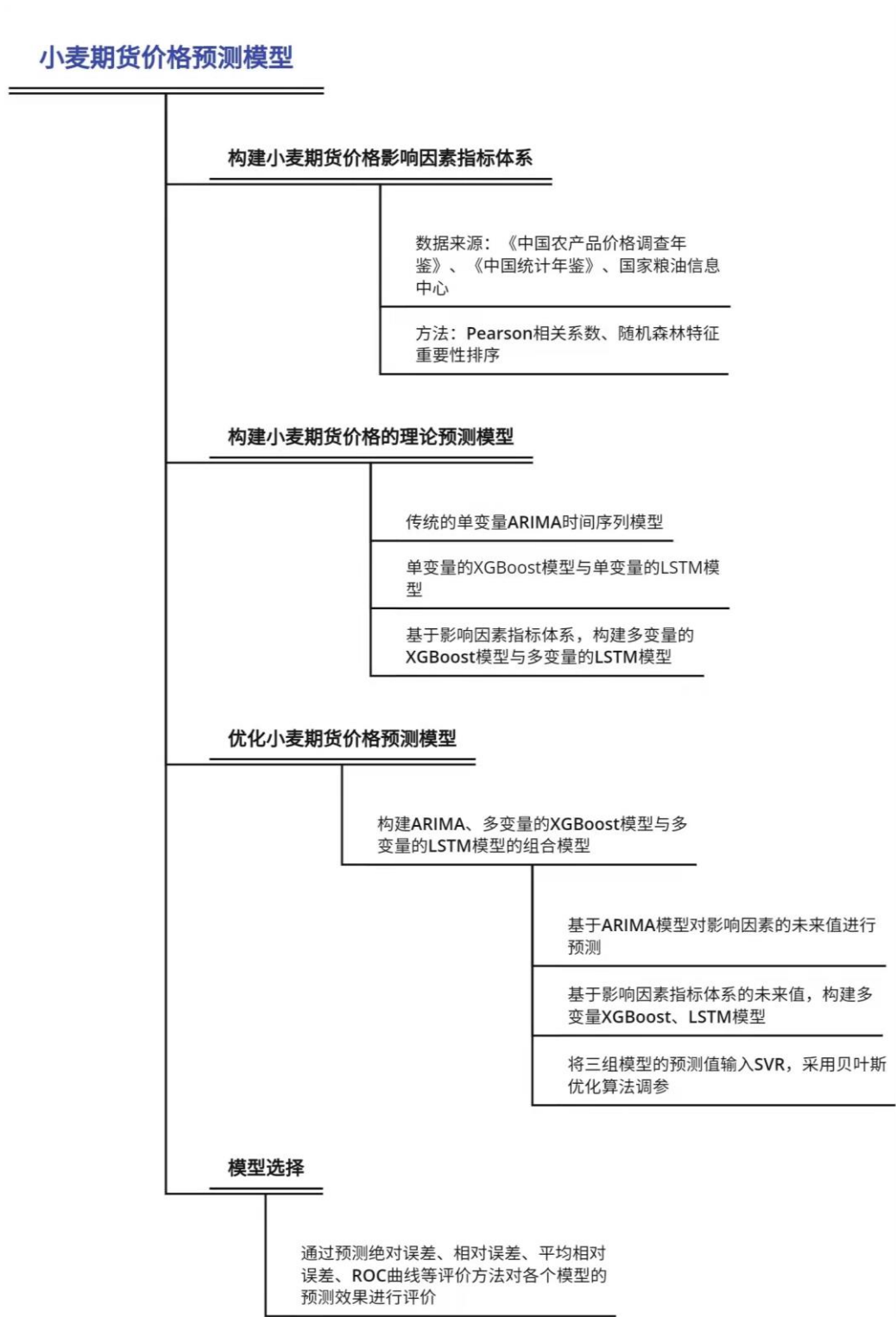


图 3-5 小麦期货价格预测模型技术路线图

3.6 人员分工

陈嘉好同学负责收集宏观粮食农业扶持改革政策、国内外政治、经济、社会影响因子的资料，确定课题研究方向，相对集中地对小麦价格变化进行定性分析。此外，对预测所得的价格走势提出未来产业布局和相关政策建议。并在论文中完成对这一部分的撰写。

苏晓钰同学负责学习 ARIMA、LSTM 等模型的原理，并对模型进行组合，构建出优化的预测模型，用 ARIMA、LSTM 等预测模型对小麦现货价格时间序列与各种变量之间的关系作出预测分析，检验比较小麦价格预测的准确率。之后，在完成前者的基础上总结几种预测模型对于小麦现货价格预测的优缺点。最后，在论文中完成各模型效用的撰写。

李敏怡同学负责查找国内外小麦期现货价格等变量与社会环境变量，通过特征提取构建出小麦现货价格影响因素指标体系，并学习掌握 R 统计分析软件的部分功能，将所构建的多个预测模型通过代码实现。最后，撰写论文的模型原理部分，并完成研究报告及相关材料的填写。

四、项目特色与创新点

(1) 选题具有现实意义

当前全球各地疫情连续不断，阻碍各国进行有规律的农业生产，而近日俄乌局势动荡，乌克兰作为欧洲粮仓的地位被打破，导致以小麦为代表的国际与国内粮食价格被不断推高，直接增加了老百姓的生活成本与市场风险，也使国家难以对小麦价格进行调控。而对小麦价格的准确预测有利于国家积极应对市场波动并采取有效的政策策略，从而降低市场风险。

(2) 研究方法具有创新性

本课题基于 stacking 继承学习算法中分类器的特点，创新性地将传统的 arima 模型以及 LSTM、XGBoost 模型组合成多变量的预测模型。目前 XGBoost 算法和 LSTM 算法在多个领域的预测问题上取得了较好的成果，但二者在经济预测方面的研究较少，结合机器学习方法在经济序列预测的较好成绩，考虑引用影响因素、使用组合模型进行预测能有效地提高模型预测效果。

五、项目进度安排

2022 年 3 月至 2022 年 5 月：梳理研究思路，撰写申报书，准备立项工作。

2022 年 5 月至 2022 年 10 月：相关资料和数据的收集、整理和分析，为课题研究准备基本的资料，研究国内小麦生产状况、农业提振政策改革及国际粮食供给趋势，参考相关领域专家意见，完成对课题所处客观环境的分析。

2022 年 7 月至 2022 年 10 月：数据处理，编写相关计算程序，建立理论模型，为课题的研究做好工具准备。

2022 年 8 月至 2022 年 11 月：测度中国各省区小麦产出和价格，并对外部环境因素进行定量分析，在此基础上分析我国小麦现货价格波动，完成对未来小麦现货价格预测，设计理论最优计量模型。在这一期间要形成初步的研究报告，撰写论文。

2022 年 11 月至 12 月：进一步向相关领域的专家咨询意见，交流信息，对研究报告及论文提出修改意见。

2023 年 1 月至 2 月：修改研究报告，完成课题，并争取在结题前发表论文 1-2 篇。

六、经费预算及依据

表 6-1 经费预算表

类 型	用 途	单 价	数 量	金 额
图书资料费	查阅国内小麦生产状况、农业提振政策改革及国际粮食供给趋势资料	200/本	20 本	4000
数据采集费	向数据库付费获取数据	100/份	20 份	2000
咨询费	咨询高校教师和相关领域专家，解决技术难点	1000 元/人	7 人	7000
劳务费	酬谢指导老师与课题组成员的付出	375 元/人	4 人	1500
印刷费	课题组内部学习资料传阅	0.5/页	1000 页	500
交通费	咨询专家产生的往返路程费用	50/次	40 次	2000
其他费用	备用资金	/	/	3000
总计	/	/	/	20000

七、项目预期成果

- （1）调研报告：通过对小麦市场公开数据建立 ARIMA-LSTM-XGBoost 预测模型分析，结合国家农产品扶持政策与小麦宏观经济环境的信息，完成一份关于小麦期货价格预测模型的报告《基于我国粮食经济发展战略背景的 ARIMA-LSTM-XGBoost 小麦价格组合预测模型》。
- （2）学术论文：将调研报告整理归纳为一篇完整的学术论文，选择合适的期刊或会议，进行投稿发表。
- （3）学术报告：向国家农业农村部递交报告，帮助相关部门通过本项目预测模型更好地预测小麦价格走势，降低市场风险。

八、申请人承诺

我保证上述填报内容的真实性。如果获得资助，我与本项目组成员将严格遵守有关规定，切实保障研究工作的进行，按时报送有关材料。

项目负责人（签名）：

苏晓钰

九、指导教师意见：

民以食为天，粮食是关乎国计民生的大事，小麦是三大主粮之一。后疫情时期政治等不确定性因素加剧，小麦价格波动剧烈，及时准确地预测小麦未来价格变化，对稳定国家粮食价格和供应具有重要意义。本项目构建了小麦期货价格影响因素指标体系，收集相关数据，比较小麦价格单一预测模型和 ARIMA-LSTM-XGBoost 组合模型，最终选出最适合的预测模型，有效预测我国小麦价格变化趋势。本项目研究选题有理论研究价值和实际意义，利用小麦价格模型有效预测的研究结论，为国家加快推进小麦及其他粮食作物定价、补贴和收储等政策改革提供重要依据。

林少萍
签 名：

十、学院（部）意见：

签名盖章：

十一、学校意见：

签名盖章：