AIGC检测·全文报告单

NO:CNKIAIGC2025FG_202505102499877　　检测时间：2025-05-13 11:15:21

篇名：多维视角下我国失业率影响因素分析与预测 模型构建——基于面板数据与机器学习方法
作者：周明慧
单位：上海对外经贸大学
文件名：新论文4.docx

全文检测结果　知网AIGC检测　https://cx.cnki.net

AI特征值：24.4%
AI特征字符数：5464
总字符数：22348

- AI特征显著（计入AI特征字符数）
- AI特征疑似（未计入AI特征字符数）
- 未标识部分

AIGC片段分布图

前部20%	中部60%	后部20%
AI特征值：5.8%	AI特征值：21.3%	AI特征值：52.6%
AI特征字符数：258	AI特征字符数：2853	AI特征字符数：2353

0　　22348

AI特征显著　AI特征疑似　未标识部分

分段检测结果

序号	AI特征值	AI特征字符数 / 章节(部分)字符数	章节(部分)名称
1	24.6%	2277 / 9239	多维视角下我国失业率影响因素分析与预测 模型构建——基于面板数据与机器学习方法_第1部分
2	10.0%	964 / 9677	多维视角下我国失业率影响因素分析与预测 模型构建——基于面板数据与机器学习方法_第2部分
3	64.8%	2223 / 3432	多维视角下我国失业率影响因素分析与预测 模型构建——基于面板数据与机器学习方法_第3部分

1. 多维视角下我国失业率影响因素分析与预测 模型构建——基于面板数据与机器学习方法_第1部分

AI特征值：24.6%　　AI特征字符数 / 章节(部分)字符数：2277 / 9239

片段指标列表

序号	片段名称	字符数	AI特征	
1	片段1	391	疑似	4.2%
2	片段2	1426	疑似	15.4%
3	片段3	258	显著	2.8%
4	片段4	462	疑似	5.0%
5	片段5	242	显著	2.6%
6	片段6	358	显著	3.9%
7	片段7	630	显著	6.8%
8	片段8	789	显著	8.5%

原文内容

摘要

就业作为民生之本，是衡量经济运行状况的核心指标。本研究基于我国31个省、直辖市、自治区2009-2021年的面板数据，选取经济、劳动力供给、产业与城镇化等维度的15个核心变量，综合运用面板数据模型与机器学习技术，从多维度系统探讨了失业率的影响机制并构建了预测模型。研究结果表明，通过固定效应模型分析发现，人口自然增长率、一般公共预算支出、外贸依存度、城镇居民家庭恩格尔系数及毕业生人数等因素对失业率具有显著影响。通过地区异质性分析进一步分析不同因素对失业率的影响差异。接着采用多元回归、随机森林和梯度提升树模型进行预测性能比较，结果显示随机森林模型在测试集上的R^2值达到0.8549，展现出最优的预测精度。基于实证研究结论，本文提出了优化财政支出结构、深化教育就业衔接机制、构建外贸内需双轮驱动体系等政策建议。

4.2%(391)

关键词：失业率；固定效应模型；随机森林；机器学习；政策建议

ABSTRACT

As the cornerstone of people's livelihood, employment serves as a pivotal indicator of economic performance. This study utilizes panel data

15.4%(1426)

from 31 provinces, municipalities, and autonomous regions in China spanning 2009–2021, selecting 15 core variables across economic, labor supply, industrial, and urbanization dimensions. By integrating panel data models and machine learning techniques, the research systematically explores the influencing mechanisms of unemployment rates and constructs prediction models from multiple perspectives. Empirical results from the fixed-effects model reveal that factors such as population natural growth rate, general public budget expenditure, foreign trade dependence, urban household Engel coefficient, and graduate numbers significantly impact unemployment. Further comparative analysis of prediction performance among multiple regression, random forest, and gradient boosting tree models demonstrates that the random forest model achieves the highest accuracy with an R^2 value of 0.8549 on the test set. Based on empirical findings, this study proposes policy recommendations including optimizing fiscal expenditure structures, deepening the education-employment linkage mechanism, and establishing a dual-driven system integrating foreign trade and domestic demand.

Keywords:Unemployment rate; Fixed-effects model; Random forest; Machine learning; Policy recommendations

1 绪论

1.1 研究背景

就业是民生的核心所在，还是衡量一个国家或者地区经济运行状况的重要指标。在我国经济做到协调发展的四个主要指标里就有充分就业这一项，其重要性十分明显。大众就业问题得到充分解决，国家经济才可能保存良好发展，社会和谐稳定才有保障，失业率高低体现就业水平，也同国家货币政策，投资者投资决策紧密相关，深入分析和准确预测失业率有益于政府有关部门执行宏观调控，改善就业政策，提升人民生活水平。

近几年，伴随产业结构的转型升级，我国的就业结构出现了大幅变化。虽然我国在经济上取得了大幅成绩，可就业问题依旧严重，失业率的起伏给社会稳定和经济发展带来了重大影响，这种情况下，探究我国失业率的影响因素及预测方法就有着重要的理论和现实价值。

从理论上讲，就业问题始终是经济学重点研究的课题，失业率属于经济波动的主要源头，其变动形式及影响要素广受关注。传统经济学理论从许多方面详细探讨了失业问题，供需关系，劳动力市场结构等，但随着大数据时代来临，机器学习技术极速发展，给失业率研究带来新思路和方法，机器学习技术可处理海量数据并挖掘背后复杂关系，为预测失业率和分析影响因素给予更有效的手段。

因此，本研究尝试找到失业率的主要影响因素以及失业率预测的最佳模型。来发现失业率变化的深层规律，给政府部门制定就业政策给予科学依照，也为未来学术上失业率的研究给予一定的参考。

1.2 研究意义

第一， 完善了失业率预测研究的指标体系，量化分析工具得到拓展。现存研究大多依靠城镇登记失业率，GDP增长率等传统指标，本研究结合相关文献，找到可能影响失业率的15个有关变量，进一步丰富了失业率预测的指标体系。

第二， 本研究利用面板数据中的固定效应模型找出了主要影响失业率的变量，比如人口自然增长率、一般公共预算支出、外贸依存度、城镇居民家庭恩格尔系数和毕业生人数等。这使得预测模型的解释性和适应性更强，给后续政策模拟供应了细致的分析工具。

第三， 本研究拓展了机器学习技术于区域就业研究里的应用范畴，近年以来，机器学习算法在处理复杂非线性关系之际彰显出突出的优越性，非常是在时间序列预测以及多变量交互分析当中冲破了传统统计方法的局限之处，经由随机森林、梯度提升树这些前沿算法，本研究试图取代传统失业率分析里线性回归模型这种单一架构，推进机器学习技术在劳动经济学领域更为深入地运用。

1.3 研究思路

本文研究的大致思路如下：

首先，收集全国31个省、直辖市、自治区2009-2021年的失业率年度数据，并结合相关文献找出与失业率可能有关的变量。其次，在数据整合完成的基础上，通过面板数据模型找到主要影响失业率的因素。最后，通过对比传统多元回归与机器学习模型的预测精度，找到失业率预测的最佳模型。

本文的研究思路见图1。

图 1 研究思路

2 文献综述

在明确了研究方向后，本章将梳理国内外关于失业率预测的研究现状。从传统的统计方法到现代的机器学习算法，为后续构建预测模型提供理论支持。

2.1 国内研究现状

我国国内关于失业率预测方法的研究主要集中在时间序列分析和计量经济模型的应用上。传统的时间序列预测方法比如ARIMA模型，被广泛用于失业率的短期预测，这个模型的优势在于对历史数据的拟合和趋势捕捉能力较强。近年来，随着机器学习技术的发展，支持向量机、随机森林等算法也被引入到失业率预测中，这些方法在处理非线性关系和复杂数据结构方面表现出色。

郑婉迪[1]（2023）针对安徽省的失业率情况展开研究。她整理1995 - 2020年统计年鉴的数据，用经济和人口改进之后的调整系数法来估算城镇失业率，以人口流向法估算乡村隐性失业率，并运用拼合预测模型，变量筛选模型和可加部分线性

2.8%(258)

模型加以分析，探究失业率及其影响因素，给就业政策的制定供应一定的参考按要求。

肖双[2]（2022）针对中国失业率预测这个问题，依靠经过变形的明尼苏达先验分布，运用贝叶斯收缩方法，创建起包含152个变量的LBVAR模型。借助比较VAR，BVAR等多种模型的预测效果，把握好从业人员指数和城镇调查失业率之间的联系，做到对中国失业率的预测，给就业工作给予参考。

原子威[3]（2021）针对我国的真实失业率情况，借助搜集1999 - 2019年《中国统计年鉴》等相关数据资料，采用ARIMA，UCSV等数种机器学习算法，并综合收缩法，因子分析法以及整合算法来形成模型加以分析，从而对我国的真实失业率实施预测，依照所得结果给出政策上的建议。

徐敏[4]（2020）考虑到我国城镇登记失业率难以确切表现实际失业情况，于是改进调整系数来考量失业率，其创建了单项预测模型，利用灰色趋势关联度等来创建定权，变权系数拼合预测模型，经过分析，得出拼合模型精度更高的结论，同时还对研究的改进方向进行了展望。

张卫泽[5]（2017）等人针对南阳市失业率预测这个问题，选取13个经济指标并以前两年失业率当作输入变量，用BP神经网络方法创建起预测模型，并借助Matlab软件做仿真实验来预测南阳市的失业率，从而给就业政策的制定给予指导。

罗圆圆[6]（2016）针对失业提示系统里指标体系创建和模型选择的问题展开探讨，她先从国民经济发展等大量方面初步选取指标，之后经由相关分析和随机森林法实施筛选，从而形成起失业提示指标体系。她运用多种模型来预测失业率，并对不同模型的效果加以比较，她还利用扩散指数判定失业率的变动情况，由此提出创建失业提示系统的思路。

2.2 国外研究现状

国外关于失业率预测方法的研究呈现出多样化和创新性的特点。传统的时间序列模型如ARIMA仍然是基础方法之一，但近年来混合模型和机器学习技术的应用日益广泛。例如，ARIMA与人工神经网络、支持向量机或自回归神经网络的结合被证明在提高预测准确性方面表现出色。此外，在机器学习领域，随机森林和遗传算法优化的神经网络等方法在处理复杂数据和捕捉非线性关系方面展现出强大的能力。深度学习技术如长短期记忆网络和门控循环单元的混合模型也被应用于失业率预测，特别是在处理时间序列数据方面表现出色。可以看出，国外研究者在不断探索如何结合更多维度的经济数据和先进的算法来提升失业率预测模型的性能和实用性。

KevinMero[23]（2024）团队着眼于厄瓜多尔的失业预测问题，提出一种名为GA - LSTM的模型，这个模型把LSTM神经网络和遗传算法融合起来，借助遗传算法去改良LSTM的参数，这样就化解了传统递归神经网络确定参数困难的问题，经过实验可知，这种混合模型在对月度失业率开展预测的时候，其性能要比BiLSTM和GRU这些传统方法好很多，从而给政策制定者给予更为精确的决策依循。

5.0%(462)

Yurtsever Mustafa[24]（2023）就美国，英国，法国和意大利的失业率预测一事，提出了一种把LSTM和GRU层融合起来的混合模型，以1983年1月到2022年5月的月度数据为依照，借助RMSE，MAPE和MAE这些指标来评定，结果找到，除了在意大利GRU模型的表现稍好一些之外，在其他三个国家，这个混合模型要比单独的LSTM或者GRU模型好得多，这就为许多国家的失业率预测赋予了一个更常见恰当的方案。

Nanik Istiyani[22]（2023）等人着眼于新冠疫情之后失业率的预测问题，运用ARIMA模型，并结合ADF验证，以2005 - 2022年的数据为按照，对2023 - 2025年的失业率加以预测。结果表明，疫情期间经济停滞社交活动受限使得失业率上升，而政府刺激政策可促使后疫情时代失业率下降，这给政策制定赋予了参考。

2.3 文献述评

国内外关于失业率预测的研究现状呈现出不同的特点和发展趋势。国内研究主要集中在时间序列分析和计量经济模型的应用上，传统的时间序列模型如ARIMA被广泛用于短期预测，近年来机器学习技术和深度学习技术也被引入，以提高预测的准确性和解释性。国内研究还注重结合区域数据（如安徽省）和宏观经济指标，为政策制定提供参考。

国外研究则更加多样化，传统时间序列模型与人工神经网络、支持向量机等混合模型的应用日益广泛，同时深度学习技术在处理复杂数据和捕捉非线性关系方面表现出色。国外研究还探索了遗传算法优化模型参数的方法，以提升预测性能。

总体来看，国内研究在方法上逐步从传统统计转向现代技术，但仍有提升空间，特别是在算法复杂性和数据维度结合方面。国外研究则更注重算法创新和理论突破，但可能在实际政策应用上不如国内研究直接。未来的研究可以进一步结合多维度经济数据和先进算法，探索更高效、实用的失业率预测模型。

2.6%(242)

3 失业问题的分析

3.1基本概念

3.1.1失业的定义

失业与就业相对，是有求职意愿劳动者无职业的状态。广义失业表现为生产资料与劳动者分离，此时劳动者难以充分发挥生产潜能与主观能动性，易造成社会资源浪费，对经济社会发展产生负面影响与破坏；狭义失业则指有劳动能力、处于法定劳动年龄阶段且有就业愿望的劳动者，失去或未获得有报酬工作岗位的现象，这即为通常所说的失业。

3.1.2失业的类型

一、自愿失业与非自愿失业

自愿失业指的是劳动者主动放弃就业机会，例如因对现行工资水平不满意或拒绝接受现有工作条件而造成的失业状态。这种失业通常源于劳动者自身的就业偏好与市场供给不匹配，往往难以通过宏观经济政策干预来解决。自愿失业的劳动者并非没有工作能力，而是基于个人选择暂时退出劳动力市场。

3.9%(358)

非自愿失业则与之相对，是指有劳动能力且愿意工作的人，在接受现行工资水平的情况下仍无法找到工作的现象。这种失业通常由宏观经济环境、产业结构调整或劳动力市场供需失衡等客观因素导致，例如经济衰退导致企业裁员、技术变革使部分技能过时、地区经济结构转型等。非自愿失业者积极寻求工作机会但受限于外部环境，是经济学研究与政策干预的重点。通常认为，非自愿失业可通过政府实施积极的就业政策、开展职业培训、优化劳动力市场结构等措施来缓解，以促进劳动力资源的有效配置。

二、摩擦性失业、结构性失业、季节性失业与周期性失业

非自愿失业又分为以下几种类型，分别是摩擦性失业、结构性失业以及周期性失业。

摩擦性失业通常是因人们更换工作等原因引致的短暂失业状况，在生产过程中难以避免。它主要源于劳动力供给层面，是一种求职型失业。企业招聘到合适员工、求职者找到心仪工作都需要耗费一定时间和精力，这主要是由于劳动力市场存在信息不对称现象。无论处于何种时期，摩擦性失业都是一种常态经济现象。

结构性失业与摩擦性失业的短期性不同，它往往是长期的。这是由于劳动力的供给和需求不匹配造成的，其典型特征是失业与职位空缺并存。这些失业者通常缺乏合适的技能，或者居住地点不适宜，无法填补现有的职位空缺。结构性失业是由经济结构变化引发的，当经济结构发生变化时，特定市场和区域中特定类型的劳动力需求低于供给，就会出现失业与职位空缺并存的情况。

季节性失业是由于消费者对某些商品和服务的季节性需求变化造成的。例如，在冬天消费者对棉衣、棉被、电暖气等商品需求旺盛；夏天则对空调、饮料、衬衫等商品需求强烈。这些需求变化通过影响某些产业的生产，进而影响对劳动力的需求，造成一定量的失业，不过这种失业是一种正常现象。

周期性失业是指当经济处于周期中的衰退或萧条阶段时，社会总需求下降所造成的失业现象。在经济衰退期，由于社会总需求不足，厂商不得不缩小生产规模，从而导致全社会普遍出现失业情况。不同行业受周期性失业的影响程度不尽相同。需求收入弹性较大的行业，在经济波动时失业影响程度较大；而需求收入弹性较小的行业，周期性失业影响则相对较小。

3.2失业的成因

一、 宏观经济环境因素

经济周期波动对就业影响显著。在经济衰退期，企业因市场需求萎缩而缩减生产规模，导致大量岗位流失，这是周期性失业的根源之一。例如制造业企业因订单锐减而裁员，直接影响相关产业工人就业。此外，通货膨胀与失业之间存在菲利普斯曲线关系，持续高通胀会干扰市场信号，影响企业投资决策，间接加大就业压力。

二、 产业结构调整因素

6.8%(630)

产业升级与技术进步带来结构性失业。劳动密集型产业向资本与技术密集型转型过程中，低技能劳动者难以适应新岗位需求。如传统纺织业自动化改造后，大量普通织布工人失业；新兴产业崛起与传统产业衰退并存，造成区域间产业转移，中西部劳动力流入地产业空心化加剧本地就业困难。

三、 教育与培训体系因素

教育资源分配不均导致技能错配。职业教育体系与市场需求脱节，职业院校课程设置滞后于产业技术迭代，学生毕业后无法满足企业对数字化、智能化岗位要求。例如人工智能领域人才缺口达500万，而每年相关专业毕业生不足30万，结构性矛盾突出。

四、 人口结构变化因素

人口老龄化加剧就业压力。一方面，老年抚养比上升使企业社保负担加重，抑制岗位创造能力；另一方面，延迟退休政策预期影响青年就业空间，新旧劳动力更替节奏紊乱。同时，高校扩招使每年应届毕业生数量接近千万，青年群体就业竞争加剧。

五、城镇化因素

快速推进的城镇化进程对我国就业格局产生了深远的影响。一方面，大量农村劳动力涌入城市，为城市的发展提供了丰富的劳动力资源，促进了城市经济的繁荣。然而，另一方面，城镇就业市场的承载能力有限，城市就业人口的快速增长使得就业竞争日益激烈。

3.3失业的影响

一、对个人的影响

失业首先会对个人的经济状况产生直接冲击，导致收入中断或大幅减少，使个人面临经济压力，难以维持原有的生活水平，甚至可能陷入贫困。同时，长期失业还会影响个人的心理健康，引发焦虑、抑郁、自尊心受损等负面情绪，削弱个人的自信心和幸福感。此外，失业也会对个人的职业发展造成阻碍，长时间脱离工作岗位可能导致技能退化，与职场脱节，降低个人在就业市场中的竞争力，影响其职业晋升和未来就业机会。

二、对家庭的影响

失业会增加家庭的经济负担，尤其是对于依赖单一收入来源的家庭，失业可能导致家庭陷入经济困境，难以支付日常开销、子女教育费用、住房贷款等必要支出。这种经济压力还可能引发家庭矛盾和冲突，影响家庭的和谐与稳定。在一些情况下，失业还可能迫使家庭成员推迟或放弃一些重要的生活计划，如购房、子女教育投资、养老规划等，对家庭的长远发展产生不利影响。

三、对社会的影响

高失业率会增加社会的不稳定因素，失业者若长期得不到有效的就业帮助和社会支持，可能会产生社会不满情绪，甚至引发社会矛盾和冲突，影响社会秩序的稳

8.5%(789)

定。同时，大量失业人口的存在也对社会保障体系构成了巨大压力，政府需要支付更多的失业救济金和其他社会保障支出，增加财政负担。此外，失业问题还可能引发一些社会问题，如犯罪率上升、贫困加剧、社会分化等，对社会的和谐发展构成威胁。

四、对经济的影响

失业意味着劳动力资源的闲置和浪费，导致经济潜在产出下降，削弱了经济的增长动力。从企业角度看，高失业率可能意味着市场需求不足，企业经营困难，进而影响企业的投资和扩张意愿，对经济的复苏和发展产生负面影响。此外，失业率上升还会进一步抑制消费市场的需求，失业者消费能力下降，导致市场需求减少，进而影响企业生产和投资，形成恶性循环，阻碍经济的持续健康发展。

4 理论基础与模型构建

4.1固定效应模型

4.1.1固定效应模型简介

固定效应模型通过控制不随时间变化的个体异质性（如地理位置、政策背景等），解决面板数据中个体效应与解释变量相关导致的内生性问题。其核心思想是引入个体或时间虚拟变量，将模型设定为：

（公式1）

其中：为被解释变量；为核心解释变量；为个体固定效应；为时间固定效应；为随机扰动项。

4.1.2固定效应模型的三种类型

一、个体固定效应模型

个体固定效应模型是面板数据分析中用于捕捉不随时间变化的个体异质性的核心方法。其核心思想是不同个体存在固有且稳定的特征差异，这些特征会影响被解释变量，但无法直接观测或量化。通过引入个体固定截距项，模型能够将这种异质性从随机误差项中分离出来，从而控制其对估计结果的干扰。

假设不同个体存在异质性截距项，模型为：

（公式2）

其中：捕捉个体特征，可通过组内变换消除。

二、时间固定效应模型

时间固定效应模型的核心思想是通过控制时间维度的系统性冲击，分离时间趋势对结果变量的影响，从而更准确地估计解释变量的因果效应。其本质是通过时间虚拟变量吸收全局时间冲击，使估计结果反映解释变量对结果的净效应。假设不同时间截距不同，模型为：

（公式3）

其中λt捕捉时间趋势（如宏观经济波动、政策变化）。

三、双向固定效应模型

双向固定效应模型通过同时控制个体与时间维度的异质性，分离不随个体变化的全局时间冲击和不随时间变化的个体固有特征，消除遗漏变量导致的内生性偏误。该模型适用于面板数据分析，能更精准估计解释变量对结果变量的净效应，常见于经济学、社会学及政策评估领域。模型为：

（公式4）

4.2多元回归模型

多元线性回归经由拟合线性方程，分析许多自变量和一个因变量的关系，用最小化均方误差法估算回归系数，常采用最小二乘法，这个模型合适于数据特征线性关系明显，样本量大的情况。多元线性回归模型可以表示为：

（公式5）

注：y是因变量，是自变量，是截距项，是回归系数，ϵ是误差项，假设服从正态分布。

计算损失函数通过最小化均方误差（MSE）来估计回归系数：

（公式6）

其中，是预测值，是真实值。

利用最小二乘法法通过求解正规方程来估计回归系数：

（公式7）

注：X是设计矩阵，包含所有自变量的观测值，y是因变量的观测值向量。

多元线性回归模型的假设条件包括：自变量与因变量之间存在线性关系。误差项之间相互独立。误差项服从正态分布。误差项的方差恒定。自变量之间不存在高度相关性。

模型构建过后通过以下指标评估模型的性能：决定系数（R^2）衡量模型解释的变异占总变异的比例。均方误差（MSE）衡量预测值与真实值之间的平均差异。均方根误差（RMSE）是MSE的平方根，与因变量的单位一致。

2. 多维视角下我国失业率影响因素分析与预测 模型构建——基于面板数据与机器学习方法_第2部分

AI特征值：10.0%　　AI特征字符数 / 章节(部分)字符数：964 / 9677

片段指标列表

序号	片段名称	字符数	AI特征	
9	片段1	349	疑似	3.6%
10	片段2	205	显著	2.1%
11	片段3	222	疑似	2.3%
12	片段4	215	疑似	2.2%

13	片段5	318	显著		3.3%
14	片段6	276	疑似		2.9%
15	片段7	240	疑似		2.5%
16	片段8	227	疑似		2.3%
17	片段9	441	显著		4.6%

原文内容

计算公式如下：

（公式8）

多元线性回归在预测和解释变量关系方面广泛应用，尤其适合数据特征之间线性关系明显且样本量较大的场景。

4.3随机森林模型

随机森林属于集成学习范畴，以构建多棵决策树并借助选举规则生成预测结果的方式来提高准确性与稳定性，每棵决策树的数据来源是随机抽取的样本子集，节点分裂依靠对特征的随机选择，以此平衡高维数据中特征数量过多或不足的问题，大幅缓解了挑选特征的压力，此外其在处理异常点表现出了较好抗干扰性，归功于众多决策树分散样本影响使得结果更均匀，赋予了模型更好的适应性和较高的执行效率去应对数据缺失或者人群分布不对等的情况，通过如此方式有助于重要变量潜在作用的浮现且拓展了实用弹性。

随机森林其核心原理是通过构建多棵决策树并集成结果以提升模型性能，适用于分类和回归任务。在分类任务中，随机森林通过双重随机性实现泛化增强：每棵决策树基于Bootstrap抽样的随机样本训练，并在节点分裂时仅使用随机特征子集，最终通过多数投票机制输出结果。回归任务则采用相似结构，但叶节点输出响应值的均值而非类别标签，通过多树平均降低过拟合风险。其底层CART决策树采用二叉树结构，以基尼指数（分类）或均方误差（回归）为分裂标准，通过特征属性将样本递归划分为互斥子集，最终在叶节点生成预测值。相较于单棵决策树，随机森林的集成策略显著提升了模型鲁棒性：通过组合多棵低相关性的弱学习器，既保留了决策树的可解释性，又克服了其易过拟合的缺陷。这种特性使其在医疗诊断、金融风控、生态数据分析等复杂场景中表现出色

3.6%(349)

随机森林主要注意以下几个关键点，基尼指数计算公式如下：

一、集成学习：随机森林通过“bagging”方法减少模型的方差，从而提高模型的稳定性和准确性。

二、基尼指数：用于决策树的节点分裂，选择最优特征和分裂点。

其中，是类别k在数据集D中的比例。

三、投票规则：随机森林通过多数投票（分类问题）或平均（回归问题）生成最终预测结果。

图 2 随机森林原理图

4.4梯度提升树模型

梯度提升机以每轮逐步推进的决策树训练为核心手段，目标就是不断削减损失函数的影响，这过程类似于一层层修正偏差现象，前序结果对后续决策树的生成起到潜移默化的作用，底层逻辑聚焦于通过压缩初期误差实现高效优化，无论是回归问题还是分类场景都可以带来不错的潜力空间，同时面对噪声环境和特异点时表现出充分韧性，替换损失函数还能适应多种复杂学习需求。其大致原理图如下：

图 3 梯度提升树原理图

模型训练部分主要包括以下4个步骤：

一、损失函数：GBM 使用损失函数衡量预测值与真实值的差异，常见的损失函数包括：

均方误差：

（公式9）

逻辑损失（Log Loss）：

（公式10）

二、梯度下降：通过计算损失函数的负梯度来指导每一步的优化：

（公式11）

其中：是第 k 步的残差。

三、基学习器：使用决策树作为基学习器，每一步训练一棵树来拟合残差：

（公式12）

四、模型更新：通过逐步叠加基学习器的预测结果：

（公式13）

其中：是学习率，用于控制每一步的更新幅度。

5 基于面板模型的我国失业率影响因素实证分析

5.1失业率数据的选取

5.1.1数据来源

本研究数据主要源自国家统计局，以确保其准确性、完整性与可靠性。国家统计局数据库提供了全国以及浙江省的部分宏观经济数据，如居民消费价格指数、人口自然增长率等，这些数据具有权威性和可比性。

2.1%(205)

5.1.2变量选取依据

要实现对城镇登记失业率的预测。首先，收集全国31个省、直辖市、自治区2009-2021年的失业率年度数据。其次，结合相关文献的研究结果，选取了15个经济指标，将这15个指标作为影响城镇登记失业率的因素。将可能影响失业率的变量按

照经济、劳动力供给、产业与城镇化三个方面进行分类，具体选择的变量见表1。变量的选取依据如下：

地区生产总值显示出一个地区的总体经济规模和发展状况，经济增长往往会带来更多就业机会，这就会对失业率造成影响。按照奥肯定律来看，经济增长和失业率呈负相关关系，人均地区生产总值更能表现居民的平均经济水准，人均GDP较高大概表明企业有更多关于增长生产和招募员工的资金，这样就会给失业率带来影响。

产业结构发生改变时，就业结构就会受到很大的影响[10]。第一产业吸纳就业的能力比较低，第一产业占比出现变动的时候，大概显示劳动力开始从第一产业过渡别的产业。第二产业可是吸纳就业的关键所在，特别是制造业，第二产业占比一旦波动起来，有关行业的就业需求就会马上被波及到。经济不断向前发展，第三产业在就业里所占的比重一点点变大，第三产业的发展情况对于吸纳就业的能力和就业的稳定来说意义独特。

城镇率上升意味着农村人口向城镇转移[11]，若城镇产业不能充分吸纳新增人口，可能导致失业率上升，但从长期看，也可能促使产业多元化发展，创造更多就业机会，降低失业率。一般公共预算支出增加，尤其投向基础设施建设、教育、医疗等领域，能直接创造大量就业岗位。

表 1 影响失业率的因素

影响因素	具体变量
经济	地区生产总值
	人均地区生产总值
	一般公共预算支出
	进出口总额
	外贸依存度
	汇率
	城镇居民人均消费性支出
	城镇居民家庭恩格尔系数
	劳动力供给人口自然增长率
	毕业生人数
产业与城镇化	第一产业增加值占GDP比重
	第二产业增加值占GDP比重
	第三产业增加值占GDP比重
	产业结构
	城镇率

汇率方面，本币升值不利于出口企业，可能使其减产裁员，导致失业率上升，而本币贬值则有利于出口企业扩大生产，增加就业机会，降低失业率。进出口总额增长通常意味着国际市场需求旺盛，国内出口企业订单增加，会扩大生产规模

，招聘更多员工，降低失业率，但进出口总额波动或下滑时，出口企业可能裁员，失业率上升。外贸依存度较高的地区或国家，经济和就业对外部市场需求依赖大，国际市场环境不好时，可能因出口受阻，企业减产裁员，导致失业率上升。

城镇居民人均消费性支出增加，表明居民消费能力强，对各类商品和服务需求增加，会刺激企业扩大生产规模，增加就业机会[12]。恩格尔系数显示居民的消费结构与生活水平状况，消费结构一旦发生改变，就会波及到相关产业的发展态势，并影响就业需求情况。

人口自然增长会对劳动力供给数量产生影响，人口自然增长率较高时可能会造成劳动力市场供过分求，进而加大失业率。毕业生属于劳动力市场的关键形成局部[13]，每年众多毕业生步入就业市场将会增多劳动力供应，进一步影响到相关行业的就业状况。

5.1.3数据预处理

在获取原始数据后，发现数据中存在一些问题，需要进行清洗处理。具体清洗操作如下：

缺失值处理：有些年份里，个别变量会出现缺失值，社会保障支出占财政支出的比重，在某些年份就没有相应的数据，针对这些缺失值，可以用多层插补法来处理，多层插补法依靠统计模型，借助多次抽样和插补去估算缺失值，这样能更精准处在理缺失数据，减小因为缺失值而产生的偏差。

异常值处理：利用箱线图法来检测数据中的异常值，比如说，找出某些年份贸易值比率数据大幅脱离正常范围的情况，也许是特别贸易政策或者骤发状况引发的。针对异常值，用四分位距法予以修正，经由计算数据的四分位数，明确异常值的界限，再把异常值换成合理数值，上下四分位数加上或者减去1.5倍IQR区间内的数值。

5.2探索性数据分析

5.2.1描述性统计分析

对收集到的数据展开阐述性统计分析，算出各个变量的均值、标准差，最小值、最大值这些统计量，就能够知晓各个变量的基本特性和分布状况。具体的变量描述性统计见表2:

表 2 变量的描述性统计

变量均值标准差最小值最大值

失业率（%） 3.29 0.65 1.21 4.61

人口自然增长率 4.757 3.133 -5.11 11.47

地区生产总值 23252.548 21074.863 445.67 125000

人均地区生产总值（元/人） 51560.312 28848.709 10814 187526

第一产业增加值占GDP比重 9.977 5.28 0.23 27.7

第二产业增加值占GDP比重 48.502 9.297 32.28 83.7

第三产业增加值占GDP比重 41.523 8.348 15.8 62

城镇率 0.572 0.137 0.223 0.938

产业结构 1.306 0.708 0.527 5.244

一般公共预算支出（亿） 4667.166 2969.809 432.36 18247.01

汇率 6.555 0.264 6.143 6.908

进出口总额（亿元） 8574.983 14265.226 21.429 82519.152

外贸依存度 0.272 0.295 0.008 1.464

城镇居民人均消费性支出（元/人） 20602.873 7597.355 8890.8 51294.6

城镇居民家庭恩格尔系数（%） 32.099 5.34 19.3 50.7

毕业生人数 22.069 14.474 0.82 67.84

5.2.2相关性分析

相关性分析是探索变量之间关系的基础方法。我们使用皮尔逊相关系数来衡量各经济指标与失业率之间的线性关系强度。相关系数范围从-1到1，其中1表示完全正相关，-1表示完全负相关，0表示无线性相关。

由表3可知，各个变量均与失业率有较强的相关关系。

表 3 相关性分析

变量相关系数 P值显著性

人口自然增长率 -0.154 0.002 ***

地区生产总值 -0.227 0.000 ***

人均地区生产总值（元/人） -0.319 0.000 ***

第一产业增加值占GDP比重 0.180 0.000 ***

第二产业增加值占GDP比重 -0.466 0.017 **

第三产业增加值占GDP比重 0.405 0.001 ***

城镇率 -0.180 0.001 ***

产业结构 -0.500 0.000 ***

一般公共预算支出（亿） -0.231 0.000 ***

汇率 0.023 0.029 **

进出口总额（亿元） -0.324 0.000 ***

外贸依存度 -0.294 0.000 ***

城镇居民人均消费性支出（元/人） -0.351 0.000 ***

城镇居民家庭恩格尔系数（%） 0.133 0.015 **

毕业生人数 0.052 0.000 ***

注：***、**、*分别代表1%、5%、10%的显著性水平

5.2.3多重共线性分析

多重共线性描述了自变量之间的关联情况，这种情况会让回归系数的估计变得不稳定，同时放大标准误差，从而削弱模型的解释能力和预测准确性，方差膨胀因

子成为衡量多重共线性的重要参考，其计算方式已被广泛采用。一般认为，当0<VIF<10时，变量间不存在多重共线性。当VIF>10时，则说明变量之间存在严重的多重共线性。

选取的初始变量的VIF检验结果见表4，可以发现有几个变量的VIF值过高，远大于10，存在多重共线性，需要进行剔除。

表 4 初始VIF检验结果

变量	VIF值
第三产业增加值占GDP比重	209.03
第二产业增加值占GDP比重	168.16
第一产业增加值占GDP比重	67.83
地区生产总值	39.69
人均地区生产总值	23.58
城镇居民人均消费性支出（元/人）	22.21
一般公共预算支出（亿）	17.64
进出口总额（亿元）	15.02
产业结构	10.16
毕业生人数	10.06
外贸依存度	9.72
城镇率	7.95
城镇居民家庭恩格尔系数(%)	2.94
人口自然增长率(‰)	1.95
汇率	1.12

因此，逐步淘汰高VIF值变量，最终存留了11个变量。经检验，这些变量的VIF值均小于10，可以进行后续的建模分析。最终保留的具体变量见表5。

表 5 剔除后的VIF检验结果

变量	VIF值
外贸依存度	8.44
一般公共预算支出（亿）	8.25
城镇率	7.83
人均GDP（元/人）	6.78
进出口总额（亿元）	6.73
毕业生人数	4.10
第一产业增加值占GDP比重	2.67
城镇居民家庭恩格尔系数(%)	2.57
产业结构	2.54
人口自然增长率(‰)	1.85

汇率 1.08

5.3基于固定效应模型的建立我国失业率影响因素分析

5.3.1模型选择

本文使用Stata软件进行建模，基于前文理论部分介绍，面板数据模型包括三种：混合回归、固定效应、随机效应。具体采用哪种模型需要通过对应的检验方法来判断。运用F检验、Hausman检验进行模型的选取。检验得到最终选择结果如表6所示。F检验与Hausman检验的p值均远小于0.01，在1%的显著性水平下拒绝原假设，因而选择固定效应模型。

表 6 模型选择结果

统计量 P值检验结果

F检验 2.41 0.000 固定效应模型

Hausman检验 43.68 0.000 固定效应模型

5.3.2基于模型结果的实证分析

基于前文模型选择的分析，将按照31个省市自治区的面板数据构建固定效应模型，带入经过VIF检验后的11个变量，得到模型的回归结果如表7所示。

从表7可以看出，有几个变量的系数并不显著，模型的回归结果不佳。

表 7 固定效应模型回归结果

变量系数 P >| t |

人口自然增长率(‰） 0.026 0.049

人均GDP（元/人） 0.007 0.789

第一产业增加值占GDP比重 0.030 0.066

城镇率 -0.447 0.609

产业结构 0.027 0.800

一般公共预算支出（亿） -0.001 0.000

汇率 0.127 0.059

进出口总额（亿元） 0.001 0.142

外贸依存度 -1.910 0.000

城镇居民家庭恩格尔系数(%） 0.025 0.004

毕业生人数 0.037 0.000

常数 1.652 0.026

对模型进行优化，最终得到5个显著影响失业率的变量，模型回归结果见表8。可以看到，各个变量系数在10%的显著性水平下均通过了显著性检验，说明这几个变量都对失业率有显著影响。

表 8 优化后的固定效应模型回归结果

变量系数 P >| t |

人口自然增长率(‰） 0.023 0.034

一般公共预算支出（亿） -0.001 0.000

外贸依存度 -2.186 0.000

城镇居民家庭恩格尔系数(%) 0.034 0.000

毕业生人数 0.033 0.000

常数 2.54 0.000

人口自然增长率的系数为0.023，表明人口自然增长率每提高1‰，失业率上升0.023%。这一正向关系可能与劳动力供给的动态变化有关：人口自然增长率的增加，意味着劳动力供给的增加，如果经济增长未能同步创造足够就业岗位，可能导致劳动力市场供大于求，进而导致失业率上升。

2.2%(215)

一般公共预算支出的系数为-0.001，说明一般公共预算支出每增加1亿元，失业率下降0.001%。尽管系数绝对值较小，但其高度显著性表明政府支出对降低失业率具有稳健的调控作用。这一结果符合凯恩斯主义理论：扩大财政支出可通过投资基础设施建设、公共服务等领域创造更多的就业机会，同时刺激社会总需求，间接带动企业用工需求，从而降低失业率。

外贸依存度的系数为-2.186，说明外贸依存度每增加1%，失业率会下降2.186%。外贸依存度反映了一个地区或国家经济对对外贸易的依赖程度，较高的外贸依存度意味着对外贸易在经济中占据重要地位。当外贸依存度增加时，出口产业繁荣，相关企业扩大生产规模，雇佣更多劳动力，创造更多的就业机会，从而降低失业率。

3.3%(318)

城镇居民家庭恩格尔系数对失业率影响极其显著且呈正相关。恩格尔系数每提高 1%，失业率随之上升 0.034 个百分点。恩格尔系数反映食品支出占比，恩格尔系数上升，意味着居民消费结构中食品支出占比增加，通常反映出居民收入水平下降，消费能力减弱。这可能暗示居民可支配收入较低，消费能力不足，进而影响企业的生产和扩张，最终导致失业率上升。

毕业生人数系数为0.033，意味着毕业生人数每增加1个单位，失业率会上升0.033个百分点。毕业生作为劳动力市场的新进入者，毕业生人数的增加直接增加了劳动力供给，如果市场上的就业机会没有相应增加，就会导致失业率上升。

2.9%(276)

5.3.3地区异质性分析

为进一步检验各因素对失业率影响的地区异质性，本文将样本划分为东部、中部和西部地区，实证结果见表9。

人口自然增长率在三大区域均对失业率产生显著负向影响，且东部地区的影响系数最大，中部和西部地区的影响系数分别为-0.088和-0.061，表明人口增长对降低东部地区失业率的作用更为突出。

一般公共预算支出在西部地区对失业率具有显著正向影响，而在东部和中部地区未达到显著性水平，说明财政支出在西部地区可能未能有效缓解失业压力，甚至

2.5%(240)

存在一定的抑制效应。

外贸依存度在东部和西部地区均对失业率产生显著正向影响，而在中部地区未显著，反映出外向型经济波动对东部和西部地区就业形势的影响更为敏感。

城镇居民家庭恩格尔系数仅在西部地区对失业率呈现显著负向影响，表明居民消费结构改善有助于西部地区失业率的降低。

毕业生人数在中部和西部地区均对失业率产生显著负向影响，而在东部地区未显著，说明人力资本积累对中西部地区失业率的缓解作用更为明显。

综上，地区间失业率影响因素存在显著异质性，东部地区在多项指标上表现出更强的就业韧性，而中西部地区则在人力资本和消费结构等方面对失业率的改善具有更为重要的作用。

表 9 地区异质性检验结果

变量东部中部西部

人口自然增长率 0.132**(p=0.043) 0.088**(p=0.015) 0.061**(p=0.034)

一般公共预算

支出 -0.001

（p=0.437） -0.002

（p=0.075） -0.001**（p=0.023）

外贸依存度 -1.494***

(p=0.002) -1.876 (p=0.189) -3.151***

(p=0.000)

城镇居民家庭

恩格尔系数 -0.865（p=0.143） 0.025

（p=0.439） 0.033**

（p=0.02）

毕业生人数 0.018

(p=0.522) 0.034***

(p=0.012) 0.061***

(p=0.006)

样本量 91 117 195

F值 27.21 7.35 17.9

6 多维视角下我国失业率预测模型的构建及分析

在对城镇失业率的影响因素进行分析后，接下来将进入失业率预测模型的构建阶段。本节将重点比较多种模型在失业率预测中的性能表现，探讨不同算法在训练集和测试集上的适用性与优劣。

本文除采用传统线性回归分析之外，还运用机器学习模型对失业率加以预测，其中有关随机森林、梯度提升树等模型。借助对比不同模型的性能，可以找出最

合适预测失业率的方法。

将数据按8：2的比例划分成训练集和检测集，用训练集来训练模型，之后在训练集和检测集上去评定模型性能。评定指标涉及R^2，MSE，RMSE，MAE以及MAPE。

6.1不同预测模型性能比较分析

6.1.1训练集模型对失业率的预测性能对比

为了评估各模型在拟合数据上的表现，本小节首先分析各模型在训练集上的性能指标，包括R^2、MSE、RMSE等，以验证模型对历史数据的拟合能力。

表 10 各模型训练集的性能指标

模型 R2 MSE RMSE MAE MAPE

多元回归 0.7691 0.0920 0.3034 0.2193 0.0721

随机森林 0.9626 0.0149 0.1221 0.0836 0.0280

梯度提升树 0.9356 0.0257 0.1602 0.1200 0.0391

从上面这个表能够看到，在训练集里，随机森林模型的表现非常好，其R^2的值达到了0.9626，这体现这个模型很适合用来拟合训练数据。

6.1.2测试集模型对失业率的预测性能对比

在完成对训练集的分析后，接下来将转向测试集的性能评估。通过对比各模型在测试集上的表现，可以进一步验证其泛化能力和对新数据的预测效果。

从表11能够看到，测试集中，随机森林模型的表现很优良，R^2的值达到0.8549，这显示这个模型有着不错的泛化能力。

表 11 各模型测试集的性能指标

模型 R2 MSE RMSE MAE MAPE

多元回归 0.7894 0.1105 0.3324 0.2566 0.0909

随机森林 0.8549 0.0761 0.2759 0.2098 0.0788

梯度提升树 0.8464 0.0806 0.2839 0.2302 0.0819

6.1.3训练集与测试集在失业率预测中的性能交叉分析

基于上述分析，本小节将综合训练集和测试集的性能指标，评估各模型是否存在过拟合现象，并筛选出最适合失业率预测的模型。通过对比模型在训练集与测试集上的表现，可评估其泛化能力及是否存在过拟合现象。

图 4 各模型在训练集和测试集上的对比

由图4可知，多元回归模型于训练集和检测集上的R^2较低，而随机森林模型在训练集与测试集表现都较好，具备良好的预测能力。

6.2失业率预测的多模型结果深度解析

在完成模型性能的总体比较后，本节将进一步深入分析各模型的具体表现，探讨其在失业率预测中的优劣势。

6.2.1多元回归模型的失业率预测结果分析

多元线性回归属于经典的统计方法，其假定因变量跟自变量存在线性关系，这

4.6%(441)

种方法简单又极易解释，不过对于非线性关系的捕捉就比较受限[8]。总体来看，多元回归模型在测试集上表现不佳，R^2值为0.7984，对失业率的预测能力较弱，不适合用作失业率的预测模型。

表 12 多元回归模型的性能指标

多元回归 R2 MSE RMSE MAE MAPE

训练集 0.7691 0.0920 0.3034 0.2193 0.0721

测试集 0.7894 0.1105 0.3324 0.2566 0.0909

图 5 多元回归训练集预测

图 6 多元回归测试集预测

6.2.2随机森林模型的失业率预测结果分析

随机森林属于一种整合学习方法，其借助形成众多决策树然后求均值的方式，来提升预测的精准度，并抑制过拟合现象。它可应对高维数据，不需要实施特征选择[7]。

表 13 随机森林模型的性能指标

随机森林 R2 MSE RMSE MAE MAPE

训练集 0.9626 0.0149 0.1221 0.0836 0.0280

测试集 0.8549 0.0761 0.2759 0.2098 0.0788

随机森林模型的性能指标见表13，可以看到，随机森林模型在训练集与测试集上的表现比较一致，其R^2分别达0.9626和0.8435，说明模型有较好的泛化能力，表明这个模型可较好地阐释失业率的变动情况，可用作失业率的预测模型。

图 7 随机森林训练集预测

图 8 随机森林测试集预测

6.2.3梯度提升树模型的失业率预测结果分析

梯度优化树属于一种融合学习方法，借助顺序塑造决策树更正前面模型所犯错误[9]。它可应对多种数据类型，预测性能往往较好。

3. 多维视角下我国失业率影响因素分析与预测 模型构建——基于面板数据与机器学习方法_第3部分

AI特征值：64.8%　　AI特征字符数 / 章节(部分)字符数：2223 / 3432

片段指标列表

序号	片段名称	字符数	AI特征	
18	片段1	228	疑似	6.6%
19	片段2	1788	显著	52.1%
20	片段3	435	显著	12.7%

原文内容

表 14 梯度提升树模型的性能指标

梯度提升树 R2 MSE RMSE MAE MAPE

训练集 0.9356 0.0257 0.1602 0.1200 0.0391

测试集 0.8464 0.0806 0.2839 0.2302 0.0819

梯度提升树模型性能指标见表14，可以发现，梯度提升树模型在训练集以及测试集上的表现比较一致，其R^2的值在训练集是0.9356，在评定集是0.8464，这显示该模型有着不错的泛化能力。

图 9 梯度提升树训练集预测

图 10 梯度提升树测试集预测

总体来讲，梯度改良树模型在测试集上的表现非常好，其R^2的值为0.8464，这表明这个模型对失业率变化的解释能力较强。

6.3结论

基于上述分析，提炼出关键结论：与传统的多元回归模型相比，机器学习模型预测精度明显提高。其中，随机森林的预测精度最高，在测试集中R^2的值达到了0.8549，这为预测失业率给予了技术支持。因此，在失业率预测模型的选择上，可以考虑建立随机森林模型来预测失业率。

6.6%(228)

7 总结与展望

7.2政策建议

基于本研究对浙江省城镇失业率影响因素的分析与预测模型的实证结果，结合当前就业市场的实际需求与政策环境[14]，提出以下政策建议，旨在通过多维度、系统化的措施优化就业结构、提升就业质量，为区域经济高质量发展提供支撑。

第一，深化教育体系与就业市场的动态衔接，缓解毕业生供需矛盾[15]。 研究显示毕业生数量是影响失业率的关键因素，反映出高等教育与市场需求存在结构性脱节。建议教育部门联合行业主管部门建立“产学研用”协同机制，推动高校专业设置与区域产业发展需求精准对接。例如，针对数字经济、智能制造等浙江省重点发展领域，增设相关专业并强化实践课程比例；同时，鼓励企业提供定向实习岗位与校企联合培养项目，缩短毕业生从校园到职场的适应周期。此外，应加强对毕业生的职业规划指导与技能培训，通过政府补贴的职业技能认证体系提升其就业竞争力，降低因技能错配导致的摩擦性失业。

第二，优化财政支出结构，强化公共政策的就业导向效应[18]。 一般公共预算支出对降低失业率具有显著作用，但需进一步聚焦就业创造效应。建议将财政资金优先投向数字经济基础设施、绿色产业园区等具有高就业乘数效应的领域，同时扩大对新业态、灵活就业的社会保障覆盖。例如，可设立专项基金支持中小微企业数

52.1%(1788)

字化转型，通过税收减免、贷款贴息等政策降低其用工成本；在公共服务领域增设公益性岗位，重点吸纳低技能劳动力与就业困难群体。此外，应建立财政支出绩效评估机制，将就业岗位创造数量纳入地方政府考核指标，确保公共资源的高效配置。

第三，构建外贸与内需双轮驱动的就业稳定机制。研究证实外贸依存度对失业率具有显著负向影响，但过度依赖国际市场易受外部冲击。建议在巩固传统外贸优势的同时，通过“数字贸易+产业集群”模式提升产业链韧性。例如，依托浙江跨境电商综试区建设，培育本土品牌出海服务平台，带动中小企业融入全球价值链；同步推进“内外贸一体化”改革，引导外贸企业拓展国内市场，开发适应内需的定制化产品。此外，应加强区域自贸协定研究，建立贸易风险预警系统，帮助企业应对外需波动，避免因订单骤减引发大规模失业。

第四，完善数据驱动的就业监测与政策模拟体系。本研究表明随机森林模型在失业率预测中具有显著优势。建议整合统计、人社、教育等多部门数据，构建省级就业大数据平台，实时监测劳动力市场供需变化；同时开发政策仿真模块，模拟最低工资调整、产业补贴等政策对就业的潜在影响[17]。例如，在预测到某行业未来半年可能出现岗位收缩时，可提前启动转岗培训计划，实现失业治理从被动应对向主动干预转变。

第五，实施差异化人口政策，缓解劳动力结构性失衡。人口自然增长率与失业率的正相关关系表明，单纯增加劳动力供给可能加剧就业压力。建议建立“人口质量红利”替代传统数量红利的政策框架：一方面，通过育儿补贴、延长产假等举措适度调控生育率，避免人口增速过快超出经济承载力；另一方面，实施“银发人力资源开发计划”，为老年群体提供灵活就业岗位与技能再培训，缓解制造业等领域的劳动力短缺问题。同时，应优化人才引进政策，重点吸引高端技术人才与紧缺工种劳动力，提升人力资本与产业升级的匹配度[16]。

第六，推动消费升级与收入分配改革，激活就业内生动力。城镇居民家庭恩格尔系数上升对失业率的正向影响，反映出居民消费能力不足制约了服务业就业扩容。建议通过收入分配改革提高中等收入群体比重，完善最低工资动态调整机制，增强居民消费信心。例如，可试点“消费券+就业联动”政策，向重点行业（如文旅、养老）定向发放消费补贴，既刺激需求又带动相关行业就业增长。同时，鼓励企业探索股权激励、利润分享等新型分配方式，使劳动者收入与企业发展更紧密挂钩，形成“收入增长—消费升级—就业扩张”的良性循环。

第七，强化区域协同与新型城镇化建设的就业吸纳功能。城镇率对失业率的复杂影响提示需避免“空心城镇化”风险。建议以都市圈建设为载体，推动产业梯度转移与人口流动协同。例如，在杭州、宁波等中心城市重点发展高新技术产业与生产性服务业，在周边县域布局配套制造业基地，形成“核心城市创新+腹地制造”的就业网络[19]。同时，加快农业转移人口市民化进程，通过保障性住房、随迁子女

教育等公共服务均等化措施，降低城镇化成本，释放新市民群体的消费与就业潜力。

7.1研究总结

本文结合国内外学者对失业率影响因素的研究，收集了全国31个省、直辖市、自治区2009-2021年的15个与失业率相关的影响变量以及失业率的年度数据，建立面板数据模型对失业率的影响因素进行研究。之后构建传统的多元回归以及随机森林、梯度提升树模型，对我国的失业率进行预测。具体内容如下：

首先，结合相关文献以及学者们的研究结果，从经济、劳动力供给、产业与城镇化三个方面进行与失业率相关的变量的选取，共找到15个相关变量。并对收集到的数据进行描述性统计分析与相关性分析。结果发现这些变量都与失业率相关，可以进行后续的分析。

其次，采用面板数据模型对我国失业率的影响因素进行分析。对已有的变量进行多重共线性检验，通过剔除VIF值较高的变量后最终确定11个有效变量，这些变量之间不存在多重共线性。通过F检验和Hausman检验选取面板数据中固定效应模型对我国失业率影响因素进行分析，最终得到5个因素在10%的显著性水平下对我国失业率均有影响，分别是人口自然增长率、一般公共预算支出、外贸依存度、城镇居民家庭恩格尔系数和毕业生人数。接下来通过地区异质性分析发现，中西部地区则在人力资本和消费结构等方面对失业率的改善具有更为重要的作用。

最后，建立传统的多元回归模型和机器学习中的随机森林、梯度提升树模型对我国失业率进行预测。通过比较不同模型在测试集与训练集上的预测性能来选择模型。结果发现，无论是R2 、MAE、MSE、RMSE还是 MAPE，随机森林的预测结果都优于其余两个模型，表明随机森林模型在预测失业率上具有更好的预测精度，是最佳的失业率预测模型。

7.3研究展望

本研究在失业率影响因素分析与预测模型选择方面取得了一定进展，但仍存在一些局限性。未来研究可从以下几个方面进行改进和拓展：

首先，扩大样本量以提高统计推断的可靠性，确保模型结果更具代表性和稳定性；其次，充分考虑时间序列特性，包括自相关和季节性因素，以更准确地捕捉失业率的动态变化规律；再次，纳入更多相关变量，减少遗漏变量偏误，从而提升模型的解释力；最后，深入探究变量间的非线性关系，以更全面地反映失业率的复杂影响机制。

通过以上改进步骤，未来研究有望进一步提升模型的预测能力与解释力，为失业率预测和政策制定提供更科学的依据。

致谢

本研究的完成离不开多方支持与帮助，在此致以诚挚谢意。

首先，衷心感谢本文的导师柯蓉教授。从选题方向到研究框架，从模型构建到

论文撰写，导师始终以严谨的治学态度和深厚的学术素养给予悉心指导。尤其在机器学习模型优化与案例分析部分，导师的真知灼见为研究突破提供了关键思路。同时，感恩家人的理解与鼓励。父母的支持是本文完成学业的动力源泉，朋友在研究瓶颈期的陪伴与督促让本人始终保持前行的勇气。

由于本人学识有限，论文仍存在诸多不足，恳请各位专家学者批评指正！

说明:

1、支持中、英文内容检测;
2、AI特征值=AI特征字符数/总字符数;
3、红色代表AI特征显著部分，计入AI特征字符数;
4、棕色代表AI特征疑似部分，未计入AI特征字符数;
5、检测结果仅供参考，最终判定是否存在学术不端行为时，需结合人工复核、机构审查以及具体学术政策的综合应用进行审慎判断。

cx.cnki.net