参赛队号：（参赛队无须填写，参赛队号由大赛官网自动生成）

2024年（第十届）全国大学生统计建模大赛参 赛 作 品

|  |  |
| --- | --- |
| 参赛学校： | 上海建桥学院 |
| 论文题目： | 《基于大数据与人工智能的长三角地区环境治理成本效益影响因素模型构建与预测》 |
| 参赛队员： | 叶子晋 姚宏伟 程铭涛 |
| 指导老师： |  |

《基于大数据与人工智能的长三角地区环境治理成本效益影响因素模型构建与预测》

方正小标宋，三号

行距：固定值24磅

黑体，四号

行距：固定值24磅

摘要

本研究旨在探索长三角地区，包括上海、江苏、浙江三省市，经济发展迅速，但同时也面临着环境治理成本效益的影响因素的挑战，并借助大数据与人工智能技术构建相应预测模型。首先，我们基于长三角地区环境治理的现状和需求，建立了环境治理成本效益指标体系，包括环境改善效益、经济成本和社会效益等方面的指标。其次，利用大数据技术，收集并分析了长三角地区的环境数据、经济数据和社会数据，为后续建模提供了充分的数据支撑。随后，运用人工智能技术，构建了环境治理成本效益影响因素的预测模型，以揭示各因素对环境治理成本效益的影响程度和趋势。最后，通过模型的验证和预测，为长三角地区的环境治理政策制定和实施提供了科学依据和决策支持。

**关键词：**长三角地区、大数据与人工智能、环境治理成本效益、影响因素、预测模型

黑体，小四号

宋体，小四号

行距：固定值24磅

目录

黑体，四号

行距：固定值24磅

[附件2 2](#_Toc165949746)

[一. 引言 7](#_Toc165949747)

[(一) 研究背景与意义 7](#_Toc165949748)

[(二) 研究背景与意义 7](#_Toc165949749)

[(三) 研究思路与方法 7](#_Toc165949750)

[(四) 创新与优点 7](#_Toc165949751)

[二. 符号说明 7](#_Toc165949752)

[三. 数据预处理 7](#_Toc165949753)

[(一) 数据来源及建模工具 7](#_Toc165949754)

[(二) 数据来源及建模工具 7](#_Toc165949755)

[1. 异常数据检测与处理 7](#_Toc165949756)

[2. 空缺数据处理 7](#_Toc165949757)

[3. 数据相关检验与可视化结果 7](#_Toc165949758)

[4. 极小化指标极大化处理 7](#_Toc165949759)

[5. 数据归一化处理 7](#_Toc165949760)

[四. 环境治理效益指数的量化研究 8](#_Toc165949761)

[(一) 指标权重赋权 8](#_Toc165949762)

[(二) 构建长三角地区环境治理效益综合评价模型 8](#_Toc165949763)

[(三) 长三角地区环境治理效益测度结果及分析 8](#_Toc165949764)

[1. 综合评价结果及分析 8](#_Toc165949765)

[2. 各指标测度结果及分析 8](#_Toc165949766)

[五. 长三角地区环境治理效益研究及实证分析 8](#_Toc165949767)

[(一) 指标的选取 8](#_Toc165949768)

[1. 内生变量 8](#_Toc165949769)

[2. 外生变量 8](#_Toc165949770)

[(二) Time-LLM模型介绍 9](#_Toc165949771)

[(三) 构建XX模型及实证结果 9](#_Toc165949772)

[1. XX参数检验 9](#_Toc165949773)

[2. XX参数选择 9](#_Toc165949774)

[3. XX参数估计结果 9](#_Toc165949775)

[六. 结论与建议 9](#_Toc165949776)

[(四) 实证结论 9](#_Toc165949777)

[1. XX差异显著 9](#_Toc165949778)

[2. XX指标各指标测度结果及分析 9](#_Toc165949779)

[3. 综合评价结果及分析 9](#_Toc165949780)

[4. 各指标测度结果及分析 9](#_Toc165949781)

[(五) 建议与对策 9](#_Toc165949782)

[1. 推动XX产业转型 9](#_Toc165949783)

[2. 调动XX主体能动性 9](#_Toc165949784)

宋体，小四号

行距：固定值24磅

说明：目录中标题需引用至正文第三级标题

表格与插图清单

黑体，四号

行距：固定值24磅

表1.XXXXXXXXXXXX

表2.XXXXXXXXXXXX

表3.XXXXXXXXXXXX

……

图1.XXXXXXXXXXXX

图2.XXXXXXXXXXXX

图3.XXXXXXXXXXXX

……

XX相对贫困治理成效的统计测度研究

宋体，小四号

行距：固定值24磅

说明：按序、依次罗列正文中所有表格、图片

# 引言

## 研究背景与意义

## 研究背景与意义

## 研究思路与方法

## 创新与优点

# 符号说明

# 数据预处理

## 数据来源及建模工具

本研究主要从江浙沪地区的主要各地级市入手进行数据调查，采用了 Python， SPSS以及 MATLAB 进行综合分析。为消除搜集数据中的主观因素影响，本研究主要通过数据网站确保其来源的真实性。本研究于以下权威数据网站进行数据调查：

各地级市的地方统计局官网

国家统计局：http://www.stats.gov.cn/

CEIC 宏观经济数据库：https://insights.ceicdata.com/

## 数据来源及建模工具

### 异常数据检测与处理

由于各项指标可能因年份原因存在一定登记性误差及对外信息披露透明度

不足问题，需对数据中异常值进行检测和处理。绿色新动能指数的子指标选取江浙沪地区主要的26个地级市 2010－2023 年面板数据，观测值有较大的时间跨度且可能存在相互关联的趋势，因此还需进行数据相关性等检验。此外，为提高数据处理精确度，本文将面板数据拆分为多个时间序列数据，针对每个时间序列数据分别进行数据的检验和处理

### 空缺数据处理

本文主要运用滑动窗口法和二次指数平滑法填补空缺数据。对于处于中间年份的空缺数据，采用滑动窗口分析法进行填补。运用多时间段信息，并把它们设定为窗口。当数据更新时，信息段的窗口将包含最新时间段的数据，保持时间段总值不变。定义数据时间序列为 ,并设定时间间隔等长。与此同时设定 为具有݊时间间隔的滑动窗口，如下图所示，当活动窗口前进到下一时间间隔时，数据将进行更新。由此填补上空缺的数据。

### 数据相关检验与可视化结果

由于涉及经济增长率、工业产值、城市化率、交通率、能源结构、月均温和降雨等与绿色发展相关的六项指标，本文通过计算六项指标之间的相关性以及对六项数据进行关联性分析。为方便说明，对相关矩阵进行可视化，最终得到以下相关性系数图：

### 极小化指标极大化处理

本文中用到的数据中同时含有极大型指标和极小型指标，在对数据进行分 析处理前，需要对不同类型的指标统进行一致化处理。在本文中，我们将非极 大型指标全部转化为极大型指标，具体公式：

### 数据归一化处理

本文中用到的数据中同时含有极大型指标和极小型指标，在对数据进行分 析处理前，需要对不同类型的指标统进行一致化处理。在本文中，我们将非极 大型指标全部转化为极大型指标，具体公式：

# 环境治理效益指数的量化研究

## 指标权重赋权

1. 评价指标：空气质量指数（AQI）、PM2.5浓度、水质指数（水质类别）、土壤污染指数。
2. 评价指标的重要性：
3. 空气质量指数（AQI）：它综合反映了空气中多种污染物的浓度水平，包括PM2.5、PM10、二氧化硫（SO2）、二氧化氮（NO2）、臭氧（O3）和一氧化碳（CO）。AQI是一个综合性的指标，能够全面地反映空气质量状况，对公众健康的影响较大。
4. PM2.5浓度：PM2.5是指直径小于或等于2.5微米的颗粒物，它们能深入肺部甚至血液，对人体健康构成严重威胁。因此，PM2.5浓度是评估空气质量的重要指标之一，尤其是在评价空气污染对健康影响时非常重要。
5. 水质指数（水质类别）：水是生命之源，水质的好坏直接关系到人体健康和生态系统的平衡。水质指数通过衡量水中的污染物种类和浓度来评价水体的健康状况。虽然其直接影响可能不如空气质量那么迅速和明显，但长期饮用或接触污染的水会对人体健康产生严重影响。
6. 土壤污染指数：土壤污染会影响农作物的生长，进而影响食品安全和人体健康。同时，土壤污染还会影响到地下水质量和生态环境。虽然土壤污染的直接影响可能不会立即显现，但其长期影响不容忽视。

3.权重赋值：根据确定的环境治理效益指数的权重，对各项指标进行加权处理，以便综合评价环境治理效益。

## 构建长三角地区环境治理效益综合评价模型

1. 确定评价模型：选择适合长三角地区环境治理效益评价的数学模型，如多元线性回归模型、模糊综合评价模型等。
2. 建立模型：根据选定的模型，建立起各项指标之间的数学关系，并确定计环境治理效益指数的方法。我们使用数据特征工程和数据集成与分析的方式，对数据进行集成，最终构建模型。
3. 模型验证：利用历史数据或模拟数据对建立的评价模型进行验证，检验模型的准确性和可靠性。对数据收集进行循环分析验证，进行多次数据影响因素分析，通过多次循环检测达到降低误差的作用。最后，通过模型的拟合程度和预测能力来验证模型的有效性。

## 长三角地区环境治理效益测度结果及分析

### 综合评价结果及分析

1. 分析评价结果：根据构建的环境治理效益评价模型，对长三角地区各地的环境治理效益进行量化测度，得出各地的环境治理效益指数。
2. 比较分析：对不同地区的环境治理效益指数进行比较分析，找出效益高低的差异性和原因。

### 各指标测度结果及分析

空气质量指数（AQI）：

PM2.5浓度：

水质指数（水质类别）：

土壤污染指数：

# 长三角地区环境治理效益研究及实证分析

## 指标的选取

### 内生变量

* 环境治理投入指数：衡量当地在环境保护上的财政投入。量化方式为来年度总支出，单位为人民币。数据来源为当地的政府预算报告。
* 空气质量指数：反映空气污染程度，评估环境政策的有效性。量化方式为基于PM2.5, PM10, NO2, SO2, CO和O3等污染物的浓度的AQI均值，单位为人民币。数据来源为国家环境保护局。
* 水质指数：评估水体污染程度和治理效果。量化方式为年度平均值的化学需氧量（COD）、生物需氧量（BOD）指标。数据来源为国家环境保护局。
* 固体废弃物指数：反映固体废弃物管理的效率和规模。量化方式为年度固体废物的总处理量（包括回收、焚烧和填埋），单位为吨。数据来源为国家环境保护局。
* 温室气体指数：评估气候政策的实施效果。量化方式为年排放总量，单位为吨，涵盖二氧化碳、甲烷等。数据来源为国家环境保护局。
* 噪音污染水平：反映当地的噪音污染状况。量化方式为年均噪声水平（分贝，dB）。数据来源为国家环境保护局。

### 外生变量

* 经济增长率：反映经济发展对环境资源的需求和压力。量化方式为。数据来源为国家统计局。
* 工业产值：衡量工业活动的规模，与环境污染密切相关。量化方式为。数据来源为国家统计局。
* 能源结构：反映能源消费对环境的影响。量化方式为。数据来源为国家统计局。
* 城市化率：城市扩张通常带来更高的能源消耗和更大的环境压力。量化方式为城市人口占总人口的比例，单位为百分比。数据来源为国家统计局。
* 年均气温和降雨量：气候变化对环境政策和资源管理的影响。量化方式为年均气温（摄氏度）、年降雨量（毫米）。数据来源为国家气象局。

## Time-LLM模型介绍

通过将大型语言模型（LLMs）重新编程，以适应时间序列预测任务。

## 构建XX模型及实证结果

### XX参数检验

### XX参数选择

### XX参数估计结果

# 结论与建议

## 实证结论

### XX差异显著

### XX指标各指标测度结果及分析

### 综合评价结果及分析

### 各指标测度结果及分析

## 建议与对策

### 推动XX产业转型

### 调动XX主体能动性

，世界各国都在为消除极端贫困而努力，并取得了一定的成绩。在全球范围内，极端贫困发生率迅速下降。据世界银行估计表明，极端贫困人口的比例从1990年的36.2%下降到2017年的9.3%。极端贫困的问题……

参照Zhou et al（2021）的研究，构建持续多维相对贫困脱/返贫指数包括以下四个步骤……

（一）单维相对贫困的识别

相对贫困发生率和相对贫困减贫成效都是一个整体的概念……个体福利用个指标来表示其水平指数，对某一区县，表示时期个体在福利指标上的取值……

**1.单维相对贫困脱贫成效/返贫现象的识别**

①利用Markov链模型思路

……其代表性元素的计算过程如下：



公式（1）

在式（1）的基础上，得到区县在福利指标上的脱贫率，……，即得到返贫率。

**公式序号**

在公式后依次标注公式序号，靠右对齐

**公式**

公式需为可编辑格式，不可用图片，居中对齐

表1 多维减贫成效测度的相对贫困维度指标

**表格标题**

字体、字号、行距同正文

按表序命名置于表格上方

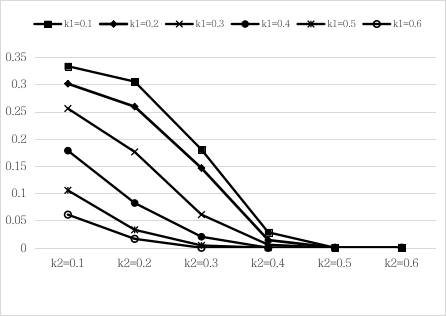
**表格内容**

宋体，小四号或五号

行距：单倍行距

注：表格应为可编辑文字，不可用图片，居中对齐

|  |  |  |  |
| --- | --- | --- | --- |
| 指标层 | 指标解释与赋值 | 临界值 | 权重 |
| 人均纯收入 | 家庭人均年纯收入（2010年不变价） | 0.4×中位数 | 1/8 |
| 身体质量指数  （BMI） | 家庭成年成员中存在BMI指数（体重（kg）/身高^2（m）） | 0.5×（中位数+18.5） | 1/8 |
| 人均商业医疗保险支出 | 上年医疗保险支出 | 0.4×平均数 | 1/8 |
| 做饭用水 | 江河湖水=1，雨水=2，窖水=3，深井水=4，自来水=5，桶装水/纯净水/过滤水=6 | 4 | 1/8 |
| 做饭燃料 | 柴草=1，煤=2，灌装煤气/液化气=3，  电=4，天然气=5，太阳能/沼气=6 | 2 | 1/8 |
| 上月伙食费 | 上月伙食费支出 | 0.4×中位数 | 1/8 |
| 上年教育支出 | 上年教育支出 | 0.4×中位数 | 1/8 |
| 存款 | 现金及存款总额 | 0.4×中位数 | 1/8 |

图1 不同临界值下我国持续多维相对贫困脱贫指数

**图标题**

字体、字号、行距同正文

按图序命名置于图下方

图居中对齐

参考文献

黑体，四号

行距：固定值24磅

1. 蒋南平, 郑万军. 中国农民工多维返贫测度问题[J]. 中国农村经济, 2017(6): 58-69.
2. 汪三贵,刘明月.从绝对贫困到相对贫困:理论关系、战略转变与政策重点[J].华南师范大学学报(社会科学版),2020(06):18-29+189.
3. 周强,张全红.中国家庭长期多维贫困状态转化及教育因素研究[J].数量经济技术经济研究, 2017, 34 (4):3-19.
4. Airio I, Moisio P, Niemelä M. Intergenerational Transmission of Poverty in Finland in the 1990s[J]. European Journal of Social Security, 2005, 7(3): 253-269.
5. 39-64.

**参考文献正文**

字体、字号、行距同正文

按先中文文献、后英文文献的顺序依次罗列参考文献

附录

黑体，四号

行距：固定值24磅

本文提及的简称与全称对应表：

|  |  |
| --- | --- |
| 全称 | 简称 |
| 持续多维相对贫困脱贫指数 | α相对脱贫指数 |
| 持续多维相对贫困脱贫成效 | α相对脱贫成效 |
| 持续多维相对贫困返贫指数 | β相对返贫指数 |

**附录内容**

字体、字号、行距同正文要求

致谢

黑体，四号

行距：固定值24磅

论文的完成离不开X老师的细心指导，从论文的选题、构思到撰写和修改……

2023年XX月XX日

**致谢内容**

字体、字号、行距同正文