Pure Mathematics 理论数学, 2022, 12(10), 1679-1692
Published Online October 2022 in Hans. http://www.hanspub.org/journal/pm
https://doi.org/10.12677/pm.2022.1210182

Hans汉斯

基于Stacking融合方法的幸福感预测与提升路径研究

张玉春

南京审计大学，江苏 南京

收稿日期：2022年9月14日；录用日期：2022年10月13日；发布日期：2022年10月20日

摘　要

幸福感作为反映居民生活水平的一个重要指标，对社会治理和社会发展具有一定的指导作用。新时代背景下，我国社会主要矛盾已经发生变化，如何更好地满足人民美好生活需要，提升居民幸福感成为政府关注的重点。本文基于中国综合社会调查(CGSS) 2015数据，通过嵌入法进行特征选择，使用逻辑回归、支持向量机、随机森林、LightGBM以及Stacking融合方法进行综合评估预测。在此基础上发现与幸福感关联度最高的三个特征并以此研究了幸福感的提升路径，对提升国民幸福感有一定参考指导意义。

关键词

幸福感，特征选择，融合方法，提升路径

Happiness Prediction and Promotion Path Based on Stacking Fusion Method

Yuchun Zhang

Nanjing Audit University, Nanjing Jiangsu

Received: Sep. 14th, 2022; accepted: Oct. 13th, 2022; published: Oct. 20th, 2022

Abstract

As an important indicator of residents' living standards, happiness plays a guiding role in social governance and social development. Under the background of the new era, the main social contradiction in China has changed. How to better meet the people's needs for a better life and improve residents' happiness has become the focus of the government. Based on the 2015 data from China Comprehensive Social Survey (CGSS), this paper conducts feature selection by embedding

文章引用: 张玉春. 基于 Stacking 融合方法的幸福感预测与提升路径研究[J]. 理论数学, 2022, 12(10): 1679-1692.
DOI: 10.12677/pm.2022.1210182

method and conducts comprehensive assessment and prediction by using logistic regression, support vector machine, random forest, LightGBM and stacking fusion method. On this basis, it finds the three features that are most closely related to happiness and studies the path to enhance happiness, which has certain reference significance for improving national happiness.

Keywords

Happiness, Feature Selection, Fusion Method, Promotion Path

Copyright © 2022 by author(s) and Hans Publishers Inc.
This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).
http://creativecommons.org/licenses/by/4.0/
Open Access

1. 引言

在过去，我国的经济发展实现了突破性增长，创造了令世界刮目相看的经济奇迹，人民的生活水平得到了极大的提高。党的十九大报告指出：“我国社会主要矛盾已经转化为人民日益增长的美好生活需要和不平衡不充分的发展之间的矛盾”。这说明，虽然人民的生活水平提高了，但人民的幸福水平没有得到同等幅度的提高。正如党的十九大报告指出：“中国共产党人的初心和使命，就是为中国人民谋幸福，为中华民族谋复兴”。提高人民幸福感已成为我们党的重要发展目标与方向。

1.1. 研究背景与意义

自 20 世纪 30 年代以来，在大众的观点里，人均 GDP 被认定为度量一个国家经济进步或国家总体实力进步的最优指标，甚至被认为是 20 世纪最伟大的发明之一。有一些国际组织提倡在发展中把提升 GDP 放在优先地位，比如世界经合组合、世界银行等。

然而，GDP 是衡量一个国家在一定时期内的总经济产出，缺乏对幸福感的感知和度量。发展不只局限于收入的增长，它是一个涉及多个维度的过程，它不仅涉及资本积累、生产和消费在数量上的增加，制度的进步、人权、民主、性别平等以及其他要素都是发展所必需的组成部分(安德鲁等，2015) [1]。因此，在评价发展状况时，应当适当提升幸福感比重，从而让评价更加客观真实，更能反映人民的生活幸福感。习近平总书记在广西考察时说，让人民生活幸福是“国之大者”。由此可见，研究如何提升与度量人民幸福感，研究如何预测人民幸福感对我国发展与人民幸福具有重要意义。

1.2. 国内外研究现状

主观幸福感早期的内涵和定义主要从情感角度来进行切入。1960 年之前，大众的普遍观点是情感的维度是一维的，且正性情感和负性情感呈现出负相关关系。Bradburn 于 1969 年首次提出了幸福程度评价的情感取向模式。Bradburn 的观点是，在情感这个维度中，正性的情感和负性的情感并不是这个维度的两个并列的存在，而是两个相互独立的情绪维度。而幸福感可以被理解为达到了正性情感和负性情感之间的平衡，正性情感会增加一个人在生活中的幸福感，相反的，负性情感会对人在生活中的幸福感产生负面影响，所以如果想要提升幸福感，就要做到减少负面情感的同时，增加正面情感(Bradburn, 1969) [2]。Diener (1984) [3]认为生活满意度和情感平衡是主观幸福感的重要组成部分。其中个体对生活的总结和认知称为生活满意度，而良好的情感平衡则是正性情感相对于负性情感占主导地位，个体在情感上具有积极主动的反应，是个体对生活中种种事件的总体性的情感感知。

对于影响幸福感的重要因素，西方学术界已经深入讨论了社会信任与国民幸福感之间的关系。Bjørnskov (1984) [4]认为社会信任与国民幸福感之间具有很强的相关关系，Delhey (2003) [5]则持有两者之间的联系并不紧密的观点。另有部分学者认为不同类型的信任与幸福感之间的关系也截然不同，如Jovanovic (2016) [6]发现，在塞尔维亚，人与人之间的信任和幸福感之间存在密切联系，而制度信任却与幸福感关系不明显。另外，部分来自中国的专家和学者也都进行了一些相关领域方面的探索和研究，袁正等(2012) [7]使用 WVS 中国的数据证明了信任对中国国民幸福感存在正向影响，后续的相关研究也都论证了这一点。值得关注的是，Kausto (2005) [8]在进行类似研究的过程中发现，虽然信任对国民幸福感有影响，但影响并不显著。在国内外，公平与幸福感之间的关系也是学者们探讨的热点之一，且重点集中在组织层面。如 Kausto (2005) [8]研究发现，在公司，具有高度组织公平感的员工，会有更加强烈的幸福感。Cassar (2015) [9]验证了，如果公司能够在程序和互动上实现公平，那么对员工的情绪会起到积极作用。赵继新和吴萌萌(2021) [10]认为组织公平对于提升员工工作幸福感具有促进作用。在非组织层面，万广华和张彤进(2021) [11]通过度量中国县区层面的机会不平等，发现机会不平等会通过影响人们社会信任或身心健康，进而对幸福感产生影响。此外，很多学者研究在收入、公平感和幸福感之间是否有密切的关系。如 Feng (2012) [12]在研究中发现，国民在居民社会保障、收入分配政策上的公平性感知对幸福感起正向影响作用。徐淑一等(2017) [13]研究发现，收入、社会阶层对幸福感的影响不及公平感对其的影响。黄嘉文(2016) [14]也在研究中发现，收入不平等会对个人体验主观幸福感起到负面作用。王洁菲和姚树洁(2022) [15]得出了收入差距的扩大会削弱城乡居民对幸福感的感知的研究结论。

1.3. 研究内容与框架

目前的国内外研究大多数集中在分析单个特征或者少数几个特征对幸福感的影响，且对影响的量化程度不够，由此可能会出现某个特征对幸福感有影响，但影响却又不显著的情况。同时，由于分析的特征数目较少，可能会忽略掉真正影响较大的特征，导致无法抓住主要矛盾，对提升国民幸福感没有明显的指导意义。在本文选择的数据集中，衡量幸福感包括性别，年龄，婚姻状况，家庭收入，社会经济地位等多种维度。同时通过数据预处理、描述性统计、特征工程、构建融合模型等多个步骤对数据集进行了详尽地分析。为了达到模型的最佳精度，本文使用了逻辑回归、支持向量机、随机森林和 LightGBM 等分类算法对模型进行训练，使用 Stacking 方法对模型进行融合，具体研究框架见图 1。通过训练模型，

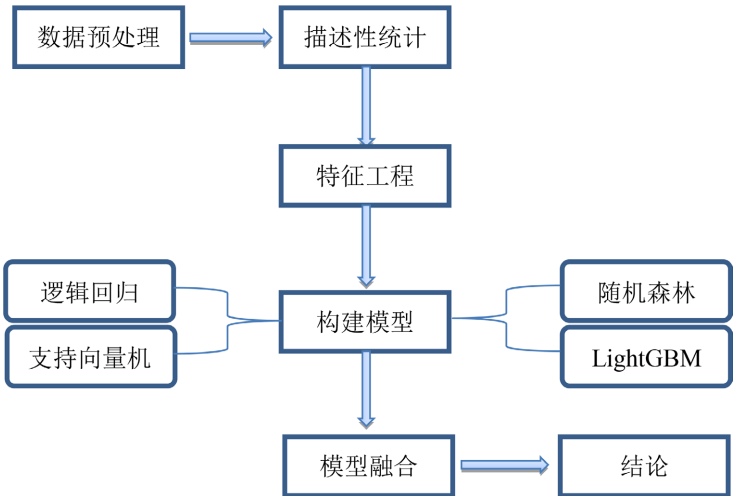

Figure 1. Research framework
图 1. 研究框架

可对幸福感进行测度，输入参数后，给出幸福或者不幸福的结论。通过特征工程，得到了各个特征对幸福感的影响程度的大小排序，清晰地展现了特征对幸福感的影响与作用，因此，从这些特征出发可以有效提升幸福感，对于政府调整政策方向，提高人民幸福感具有一定的指导意义。

本文的结构安排如下图 1：第二部分介绍数据的来源与预处理过程，第三部分对预处理后的数据进行描述性统计，第四部分使用嵌入法对数据进行特征选择，第五部分是模型融合的具体过程，第六部分是幸福感提升路径、结论与展望。

2. 数据来源和预处理

2.1. 数据来源与说明

数据来自 2015 年中国人民大学中国调查与数据中心主持的《中国综合社会调查(CGSS)》项目，该数据通过多阶段分层抽样的问卷采访获得，共有 8000 条数据。数据可在 https://tianchi.aliyun.com/dataset/dataDetail?dataId=89421 下载。该数据具有 40 个特征，包括性别、年龄、民族、宗教信仰、最高教育程度、个人收入、政治面貌、身体健康状况、家庭总收入、婚姻情况、是否有汽车，有几处房产、是否经常在空闲时间社交、是否经常在空闲时间休息放松等。研究对象为幸福感程度，包含非常不幸福、比较不幸福、不确定、比较幸福、非常幸福 5 个标签。

数据特征大多数都为分类变量，少数数值型变量，还有难以进行分析的时间格式数据，所以需要通过数据清洗、特征转化、分箱等方法进行预处理，数据预处理的质量会直接决定数据分析与预测的准确性。

2.2. 数据预处理

2.2.1. 缺失值的处理

首先利用 pandas 的 isnull 函数找寻缺失值，发现“work_status”、“work_yr”、“work_type”、“work_manage”这四个特征各有 5049 条缺失值，已超过了样本量的一半，为了不影响后续的分析，删除这四个特征；另外“family_income”有 1 条缺失值，下一步进行填充。

另外，数据集有一些包含特殊意义的值：−1 表示不适用，−2 表示不知道，−3 代表拒绝回答，−8 是无法回答。将这些值和空值一起视为缺失值处理：数值型数据的缺失值填补为均值、分类数据的缺失值填补为众数或中间态度。将缺失值填充为平均值的有：income、family_income；将缺失值填充为众数的有：religion、religion_freq、edu、political、socialize、relax、learn、family_m、family_status、house、car、status_peer、status_3_before、view、inc_ability；将缺失值填充为中间态度的有：happiness、nationality、health、health_problem、depression、equity、Class (为了和 Python 的 class 类区分，将其改名为 Class)。

2.2.2. 检查异常值

剔除数据集中背离逻辑的异常值，设置以下三个条件并剔除 854 条异常值：

条件 1：家庭人数(family_m)大于 30；

条件 2：房产数(house)大于 30；

条件 3：个人收入(income)大于家庭总收入(family_income)。

2.2.3. 时间格式和体重数据的转化

新增特征年龄(age)：由于 survey_time 问卷调查时间均为 2015 年，所以计算方法为 age = 2015 − birth，然后删除 survey_time 和 birth 这两个特征。

新增特征身体质量指数(BMI)：计算方法为

$$\mathrm{BMI}=\frac{\mathrm{weight}/2}{\mathrm{height}/100}.$$

之后删除 height 和 weight 这两个特征。

2.2.4. 特征处理

根据虚拟变量的原则，把两类分别表示为 0 和 1，将 car、survey_type 和 gender 进行转化，统一把 2 替换成 0。同时，对类别过多过复杂的分类变量进行简化，例如将 nationality 简化为 0 汉族和 1 少数民族；edu 简化为 1 = 小学、2 = 初中、3 = 高中、4 = 大学、5 = 研究生及以上，其他 hukou、marital、religion_freq、work_exper、house、family_m 也如此。最后，对数值型数据进行分箱处理。将 floor_area 分为：1 = 小面积、2 = 中等面积和 3 = 大面积；BMI 分为：1 = 瘦弱、2 = 正常、3 = 偏胖、4 = 肥胖；income 和 family_income 均分为 5 类：1 = 贫穷、2 = 低收入、3 = 平均收入、4 = 中高收入和 5 = 高收入；age 分为 5 类：1 = 35 岁及以下、2 = 36~45 岁、3 = 46~55 岁、4 = 56~65 岁、5 = 66 岁及以上。经过以上步骤，完成数据的预处理工作。

3. 描述性统计

3.1. 整体幸福感描述

在幸福感数据集中，1 代表非常不幸福，2 代表比较不幸福，3 代表不确定、4 代表比较幸福、5 代表非常幸福，即数值越大，幸福感越高。本文首先讨论整体上的幸福感状况：

Table 1. Statistical description of well-being
表 1. 幸福感的统计描述

统计指标	取值
样本总量	7146
均值	3.87
标准差	0.81
最小值	1
25%分位数	4
50%分位数	4
75%分位数	4
最大值	5

从表 1 可知，幸福感的均值为 3.87，25%分位数是 4，这表明超过四分之三的受访者具有较高的幸福感。

为了更加清晰和直观，绘出五类受访者所占比例的饼状图，如图 2 所示。由图可知，比较幸福占比 60.41%，非常幸福占比 17.73%，这说明大多数受访者处于幸福的生活状态。

3.2. 单特征统计描述

由于数据包含的特征较多，本部分仅以受教育程度和空闲时间社交的频繁程度两个特征为例，进行描述性统计分析，探究其与幸福感之间的关联度。

3.2.1. 受教育程度

首先，绘制受教育程度与幸福感程度的条形图进行，如图 3 所示。由图 3 直观可看，在评价比较幸

福的人群中，比例随着受教育程度上升而上升，在评价非常不幸福、比较不幸福、不能确定人群中，比例随着受教育程度上升而下降，仅在评价非常幸福人群中，比例随着受教育程度上升呈现出先下降后上升的趋势，但占比最高的仍是研究生及以上受教育程度。

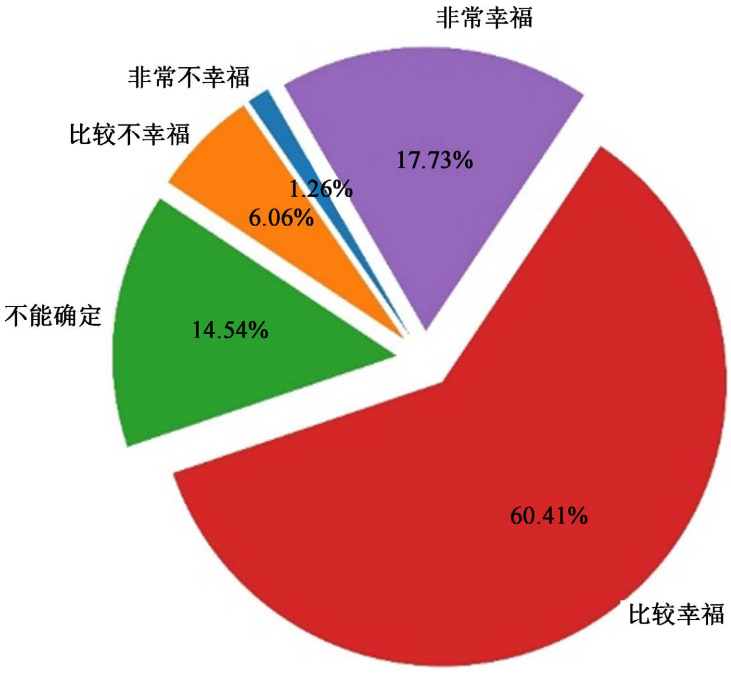

Figure 2. Happiness pie chart
图 2. 幸福感饼图

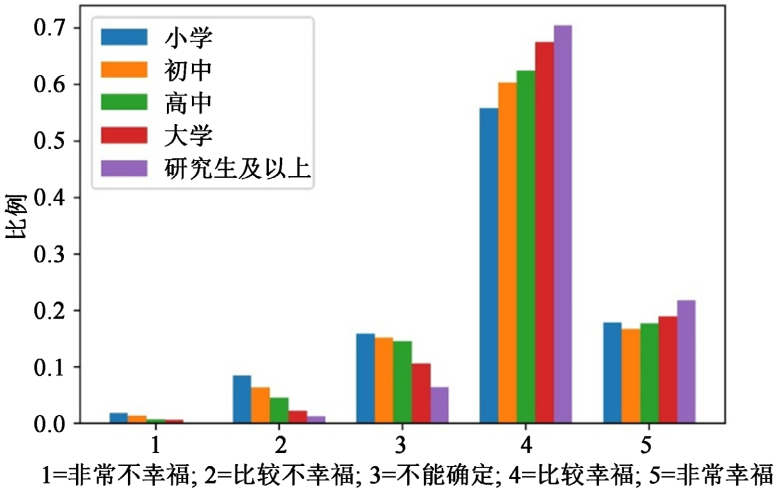

Figure 3. Bar graph of educational attainment and happiness
图 3. 受教育程度与幸福感程度的条形图

为了更全面地体现数据趋势，本文绘制点图进行分析，见图 4。由图可看知，幸福感均值随着学历上升而上升，且小学、初中、高中、大学的误差线差别不明显且都比较小，说明这四种学历在幸福感体验上比较稳定，变异程度低，而研究生及以上学历误差线较大，说明研究生及以上学历在幸福感体验上不是很稳定，不确定程度较大。图 3 和图 4 均说明受教育程度对幸福感有正向影响。

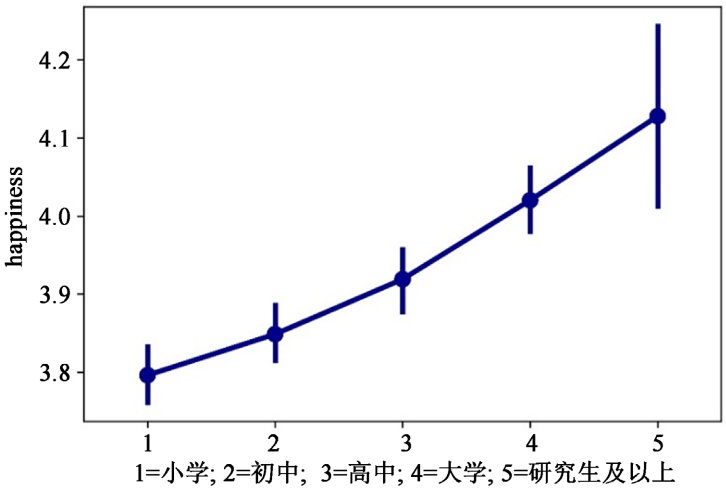

Figure 4. Dot plot of education and happiness
图 4. 受教育程度与幸福感程度的点图

3.2.2. 空闲时间社交的频繁程度

本文绘制了社交频繁程度和幸福感程度的频率表、社交频繁程度和幸福感程度的点图进行分析。在数据集中，1 代表从不社交，2 代表很少社交，3 代表有时社交，4 代表经常社交；5 代表非常频繁社交。

Table 2. Frequency distribution of social frequency and happiness
表 2. 社交频繁程度和幸福感程度的频率分布表

	从不	很少	有时	经常	非常频繁
非常不幸福	0.026798	0.014367	0.008957	0.008838	0.012384
比较不幸福	0.069111	0.071833	0.060457	0.043561	0.046440
不确定	0.143865	0.163692	0.167488	0.099116	0.092879
比较幸福	0.569817	0.583370	0.610390	0.652778	0.544892
非常幸福	0.190409	0.166739	0.152709	0.195707	0.303406

由表 2 可知，非常不幸福占比最高的人群是从不社交人群，占比为 2.68%，非常幸福占比最高的人群是非常频繁社交人群，占比为 30.34%。在经常社交人群中，评价比较幸福和非常幸福总占比 84.84%，在频繁社交的人群中，评价比较幸福和非常幸福总占比 84.82%，两者差别较小。值得注意的是，五类人群中，比较幸福占比最小是非常频繁社交人群，为 54.49%，比从不社交人群的比较幸福占比还要小。

非常频繁社交人群中非常幸福占比最高而比较幸福占比却最低，这体现了在幸福感体验上的不稳定性，因此本文绘制能体现不稳定性的点图进行分析，如图 5 所示。由图可知，非常频繁社交人群的误差线最大，说明频繁社交在幸福感体验上有一定的变异程度，较不稳定，这为频繁社交人群中评价比较幸福占比最小提供了一种解释。同时可以看出从不社交，很少社交，有时社交的幸福感均值差别不大，但从不社交具有较大的误差线，说明从不社交在幸福感体验上并不稳定，对于有些人，从不社交能带来一定幸福感。而经常社交和频繁社交在幸福感均值上得到了很大的提升，说明社交对于提升幸福感有较大的帮助。

4. 特征工程

特征选择是从原始特征中选取较为重要的特征，通过减少特征数量，达到提升模型精度和降低运算

成本的效果。本文使用基于树模型的嵌入法进行特征选择。嵌入法需要使用机器学习的某些算法来进行训练，得到各个特征的系数，特征系数越大说明特征越重要，然后确定一个最佳的阈值，大于这个阈值的特征保留，小于的删除。通常有基于惩罚的嵌入法以及基于树模型的嵌入法，本文选择极端随机树作为基模型进行特征选择。首先建立一个极端随机树模型，并求出该模型能达到的最大系数，从 0 至最大系数进行遍历循环，得出最佳阈值为 0.01934，此时能达到模型的最大精度，详细过程见图 6。

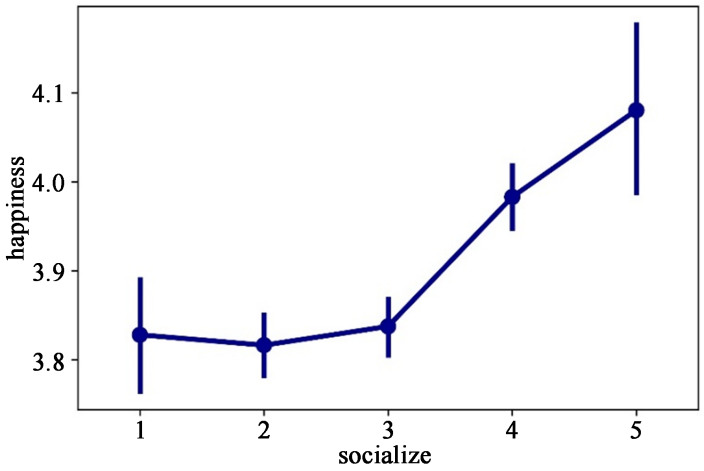

Figure 5. Dot plot of social frequency and happiness
图 5. 社交的频繁程度和幸福感程度的点图

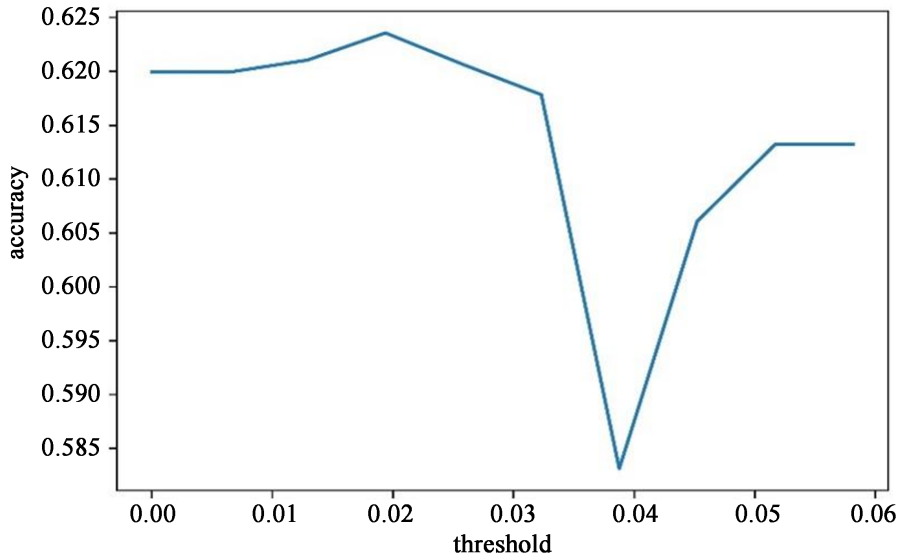

Figure 6. The variation of model accuracy with threshold
图 6. 模型精度随阈值的变化图

求出最佳阈值后，配合 SelectFromModel 函数即可获得最终的特征。通过特征选择，共保留 26 个特征。按照特征系数从大到小的顺序对特征进行排序，排序结果见表 3。

特征系数越大，说明该特征越重要，与幸福感的关联度也越高。由表 3 可知，社会公平程度、社会阶层和心情沮丧的频繁程度是与幸福感关联程度最高的三个特征，而户口类型、拥有的房产套数、性别等特征则与幸福感关联度较小。

Table 3. Ranking results of characteristic coefficients
表 3. 特征系数排序结果

特征名称	特征说明
认为当今社会公不公平	1 = 很不公平；2 = 较不公平；3 = 一般；4 = 较公平；5 = 很公平
认为自己处于哪个阶层	1 = 社会最底层；2；3；4；……；10 = 社会最顶层
心情沮丧的频繁程度	1 = 总是；2 = 经常；3 = 有时；4 = 很少；5 = 从不
社交的频繁程度	1 = 从不；2 = 很少；3 = 有时；4 = 经常；5 = 非常频繁
休息放松的频繁程度	1 = 从不；2 = 很少；3 = 有时；4 = 经常；5 = 非常频繁
健康状况	1 = 不健康；2 = 较不健康；3 = 一般；4 = 较健康；5 = 健康
年龄段	1 = 18~35；2 = 36~45；3 = 46~55；4 = 56~65；5 = 66 岁及以上
家庭年总收入	1 = 贫穷；2 = 低收入；3 = 平均收入；4 = 中高收入；5 = 高收入
观点和大众的一致程度	1 = 很不一致；2 = 较不一致；3 = 一般；4 = 较一致；5 = 很一致
个人年收入	1 = 贫困；2 = 低收入；3 = 平均收入；4 = 中高收入；5 = 高收入
健康问题的频繁程度	1 = 非常频繁；2 = 比较频繁；3 = 有时；4 = 比较少；5 = 非常少
认为目前的收入是否合理	1 = 非常合理；2 = 合理；3 = 不合理；4 = 非常不合理
身体质量指数	1 = 瘦弱；2 = 正常；3 = 偏重；4 = 肥胖
受教育程度	1 = 小学；2 = 初中；3 = 高中；4 = 大学；5 = 研究生及以上
住房面积	1 = 小面积；2 = 中等面积；3 = 大面积
学习充电的频繁程度	1 = 从不；2 = 很少；3 = 有时；4 = 经常；5 = 非常频繁
家庭人数	1 = 2 人及 2 人以下；2 = 3 人；3 = 4 人及 4 人以上
家庭经济状况	1 = 远低于平均水平；2 = 低于平均水平；3 = 平均水平
社会地位和三年前相比	1 = 上升了；2 = 差不多；3 = 下降了
工作经历及状况	1 = 非务农工作；2 = 务农；3 = 无工作
婚姻状况	1 = 未婚；2 = 已婚；3 = 离异；4 = 丧偶
与同龄人相比的社会地位	1 = 较高；2 = 差不多；3 = 较低
性别	0 = 女；1 = 男
拥有的房产套数	0 = 无房；1 = 1 套；2 = 两套和以上
户口登记状况	1 = 农业户口；2 = 非农业户口；3 = 其他
样本类型	0 = 农村；1 = 城市

5. 模型融合

在特征选择后，本文将对特征选择后的数据集使用逻辑回归、支持向量机、LightGBM 和随机森林四种常用的机器学习分类算法进行单分类器构建，再从中选取较优的分类模型进行模型融合。模型融合是融合多个不同模型的过程，和 bagging、boosting 方法的区别在于融合方法能融合不同种类的模型，而后者是集成相同种类的模型。一般情况下，多模型融合的预测精度优于单模型。

5.1. 单分类器构建

5.1.1. 逻辑回归

逻辑回归是一种分类算法，主要思想如下：寻找合适的分类函数，输入数据后，通过函数值得到分类结果；构造损失函数，用来表示预测类别和真实类别之间的差距；最小化损失函数，从而获得模型参数的最优估计。本文使用 sklearn 库的 LogisticRegression 函数构建逻辑回归分类模型，在模型优化过程中，调参是重要的一环，结合本数据实际，可得出重要参数和取值范围，见表 4。

Table 4. Logistic regression parameters and their meanings
表 4. 逻辑回归参数及含义

参数名	参数含义	本数据取值
Penalty	正则化类型	L1, L2
C	正则化系数	0.2~1
solver	优化算法	newton-cg, lbfgs, saga, sag
max iter	最大迭代次数	1~100

正则化项即罚函数，该项对模型向量进行“惩罚”，从而避免单纯最小二乘问题的过拟合问题。在模型损失函数上添加 L1、L2 范数惩罚用于防止过拟合，提升模型的泛化能力。使用 L1 可以得到稀疏的权值；使用 L2 可以得到平滑的权值。惩罚项系数 C 越大对模型的惩罚力度越大。

使用梯度下降法求解模型参数，当模型迭代一定次数之后，模型表现不再有过大变化，达到收敛，因此最大迭代次数不应过大，浪费计算资源。

在调参过程中首先对两种正则项进行选择。图 7 为模型在 L1、L2 两种正则方式下的表现，可看出使用 L2 正则项使模型在测试集上的表现更好。

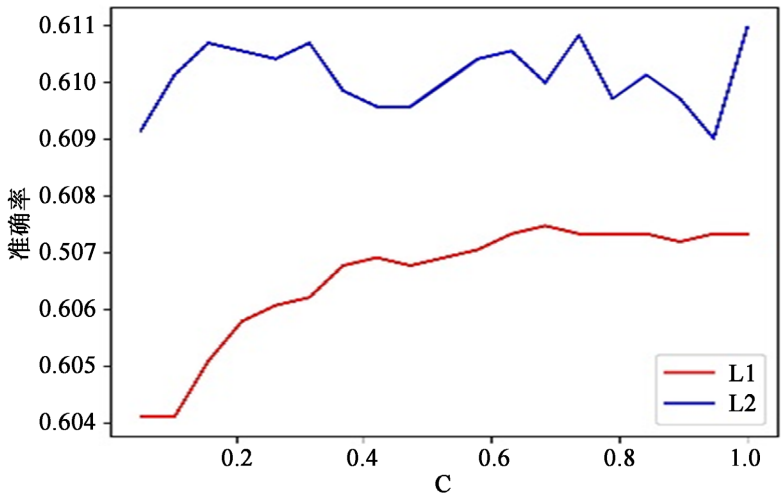

Figure 7. Logistic model performance under L1 and L2 regularity
图 7. L1、L2 正则下 Logistic 模型表现

选定惩罚项为 L2 正则并对惩罚项参数进行更细致的调整。如图 8(左)所示当惩罚项参数为 0.78 时模型准确率最高。确定正则化方式及惩罚项参数后寻找最佳梯度下降法迭代次数，如图 8(右)所示，当迭代次数为 100 此时模型预测准确率达到最大。

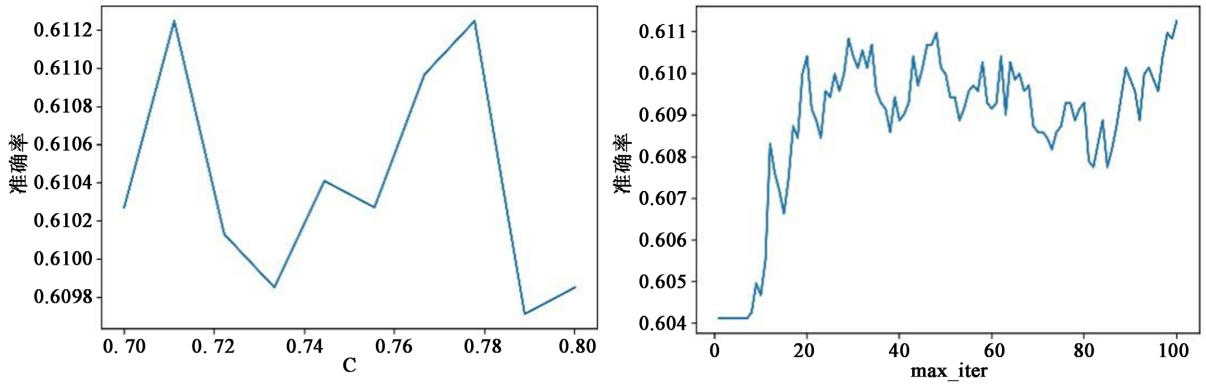

Figure 8. The influence of the penalty term coefficient and the number of iterations on the model under the L2 regularity
图 8. L2 正则下惩罚项系数及迭代次数对模型的影响

分别使用 newton-cg、lbfgs、saga、sag 迭代方法对模型做最后的调整。在确定惩罚项系数为 0.78 的 L2 正则项，最大迭代次数为 100 条件下，newton-cg 迭代方法使得模型准确率达到最大，准确率为 0.61363。以上是逻辑回归的调参过程。

5.1.2. 调参结果汇总

支持向量机、随机森林和 LightGBM 的调参过程类似于逻辑回归，这里不再赘述。在最优参数下，支持向量机的准确率为 0.62343，随机森林的准确率为 0.62981，LightGBM 的准确率为 0.62421，随机森林方法在特征选择后的数据集上具有最佳分类效果。

5.2. 融合模型构建

模型融合方法有平均法 Averaging、投票法 Voting、堆叠法 Stacking 等方法。在模型融合的多种方法中，平均法适合分析回归数据，投票法需要给各个模型设定权重，但权重设定较为主观，没有统一有效的标准，而堆叠法集成了不同的算法，各种算法取长补短，具有其他方法不具备的优势。因此，本文采用更具优势的堆叠法 Stacking 来进行模型融合。

对于两层的 Stacking 模型，原理是把原始训练集分为两部分，即训练集和验证集：在第一层模型中使用 k 折交叉验证的方法对训练集进行切分，并用训练集训练，用验证集预测；接着在第二层模型中利用验证集预测结果继续训练模型。

Stacking 方法步骤：

1) 将数据分为训练集和测试集，对训练集划分为 k 个大小基本一致的集合，取一份作为验证集，其余的为训练集；

2) 在第一层中创建多个模型，既能是相同类型的也能是不同类型的；

3) 对于每一个模型用各自的训练集训练各自的模型，训练好后对各自的验证集和测试集进行预测，得到 val_pre 和 test_pre；

4) 创建第二层模型，将每个模型对应的 val_pre 拼接作为第二层的训练集，将所有模型的 test_pre 取均值作为第二层的测试集；

5) 用第 4 步训练好的模型对第二层的测试集进行预测，得出最终的结果。

本文共选用 4 个分类模型，从分类准确率来看，逻辑回归的准确率最低，支持向量机的准确率相较于随机森林、LightGBM 也较差，因此将随机森林和 LightGBM 这两个分类准确率较高的模型放入第一层，第二层使用逻辑回归以防过拟合，以此构成融合模型。

5.3. 结果汇总

本文共有四个单分类器模型和一个融合模型，在最优参数下，最终的分类准确率汇总见表 5。

Table 5. Accuracy comparison between fusion model and single classifier model
表 5. 融合模型与单分类器模型的准确率比较

	随机森林	LightGBM	逻辑回归	支持向量机	Stacking
模型准确率	0.62981	0.62421	0.61363	0.62343	0.64028

由表 5 可知，Stacking 融合模型的分类准确率最高。Stacking 方法利用了交叉验证的思路，对数据的利用率高，因此结果也会更为稳健。在给定 26 个筛选出的特征的条件下，利用本文提出的 Stacking 融合模型，能够以 0.64028 的准确率实现对非常不幸福、比较不幸福、不确定、比较幸福、非常幸福这五个类别的分类。

6. 幸福感提升路径、结论与展望

6.1. 幸福感提升路径

由特征工程结果可知，社会公平程度、社会阶层和心情沮丧的频繁程度是影响幸福感的三个最重要因素，本文从这三个影响因素出发展开幸福感提升路径的研究。

6.1.1. 创建公平社会环境，提高民众公平感知

社会公平感对幸福感起到正向影响作用，幸福感会随着社会公平感的提升而得到提升。对于个体而言，最简单的公平就是付出了应该付出的努力后，能够得到应得的回报与利益。如果一个人没有付出应该付出的努力，没有承担着自己应该承担的责任，却比别人获得了更多的利益，就会让人感到不公平感。个体在构建未来的过程中，往往以能看到、能接触到的身边人的生活为基准，进而发现自己与他人的差距，产生了落差感与不公平感，由此降低了幸福感。因此，在相互比较中产生的社会公平感会对个体幸福感更容易产生影响。要提升我国国民幸福感，可以从提高社会公平感入手，完善收入分配制度，缩小收入分配差距，使分配更加公平公正、提供平等的公共服务资源、建立公平公正的制度、完善相关法制保障，以营造风清气正，和谐公平的社会环境。社会公平感提升了，社会各个阶层和群体的矛盾也会得到缓解，整个社会能够良性流动与良性竞争，社会氛围会变得更和谐，更温暖，国民的幸福感也会得到提升。

6.1.2. 关注低阶层民众，提高主观阶层感知

社会阶层和幸福感同样是正相关的。社会阶层是指由具有相同或相似资源、声望的社会成员组成的群体。社会阶层分为主观社会阶层与客观社会阶层。主观社会阶层强调个体的主观认知，即自己认为自己在社会里拥有怎样的资源或声望，认为自己处于什么高度和位置。客观社会阶层则根据个体的财富水平、学历程度、职业等因素来确定，更具客观性。本文所采用的数据集是问卷形式，得到的社会阶层结论是主观的。从数据集可以看出，不同社会阶层个体所感受到的幸福感具有较大差异，在主观层面上，如果个体认为自己处于较高社会阶层，就会对生活和社会具有较高的满意度，所感受到的幸福感也会更强烈，而认为自己处于较低社会阶层的个体，对生活的满意度也会较低，幸福感也会随之下降。在情感层面上，低阶层个体相较于高阶层个体，更容易产生负面情绪，更容易体会到悲伤、焦虑与压力。因此，政府在社会治理的过程中可以侧重于考虑低阶层民众的幸福感，通过提高其对国家建设和发展成果的共享感来提高主观阶层感知，进而让低阶层民众更有安全感，从而保障人民的幸福感。

6.1.3. 个体要善于调节自己，积极面对生活

心情沮丧的频繁程度对幸福感起负面作用，即沮丧越频繁越不幸福。现代社会经济运转节奏比较快，人们在生活中不可避免地面临各种各样的压力，比如受到上级责备、加班较多、学习工作压力大等等。再加上信息时代来临，所能接触到的各种各样的信息呈爆发式增长，在浩瀚的“信息汪洋”中，不可避免地存在着一些垃圾信息或者负面信息，而这些信息会使得心情沮丧的频繁程度产生一定幅度的增长。心情沮丧越频繁，对幸福感的冲击也会越大。作为个体而言，为了降低心情沮丧的频繁程度，要善于调节自己，改变自己的认知方式，不能对自己“以偏概全”，不能因为一次失败就否定自己，认为自己一无是处，要善于发现自己的优点；可以多出去社交，多和朋友倾诉谈心，用多种方式去纾解心中的压力和痛苦，努力去除负面情绪，提升自己的幸福感，让自己更积极阳光的面对生活。

6.2. 结论与展望

本文通过数据清洗、特征构造、连续性特征分箱等方法进行数据预处理；对预处理后的数据集进行了统计描述；基于嵌入法进行特征选择，由特征选择得到了特征的重要性排序并以三个最重要特征为基础进行了幸福感提升路径研究；对特征选择后的数据集使用逻辑回归、支持向量机、随机森林、LightGBM构建单分类器模型；选择分类精度较高的单分类器模型使用 Stacking 方法进行模型融合，最终的分类准确率为 0.64028，此融合模型为预测我国国民的幸福感提供了新的思路和解决路径，只要给定居民的各个特征要素，就可以预测是否幸福，为进一步探索幸福感影响因素以及提升国民幸福感提供了参考。中国是一个国土辽阔的国家，且不同地区在经济水平、风俗文化、气候地貌等方面均具有较大差异，这就导致不同地区的居民对幸福的定义和理解可能不尽相同，因此，为了更好地研究居民幸福感，后续可以分地区进行研究，对每个地区的数据单独进行分析整理，研究出每个地区的幸福感影响因素，并构建相应的幸福感预测模型，从而更有针对性地研究幸福感提升路径，给出更为精确的政策建议。

基金项目

江苏省研究生科研与实践创新计划(KYCX22_2209)。

参考文献

[1] 安德鲁·E.克拉克, 克劳迪亚·塞尼克, 肖辉. GDP 增长能否提升发展中国家的国民幸福感? [J]. 国外理论动态, 2015(12): 93-104.

[2] Bradburn, N.M. (1969) The Structure of Psychological Well-Being. *Social Service Review*, **44**, 139-157. https://doi.org/10.1086/642592

[3] Diener, E. (1984) Subjective Well Being. *Psychological Bulletin*, **95**, 542-575. https://doi.org/10.1037/0033-2909.95.3.542

[4] Bjørnskov, C. (2008) Social Capital and Happiness in the United States. *Applied Research in Quality of Life*, **3**, 43-62. https://doi.org/10.1007/s11482-008-9046-6

[5] Delhey, J. and Newton, K. (2003) Who Trusts? The Origins of Social Trust in Seven Societies. *European Societies*, **5**, 93-137. https://doi.org/10.1080/1461669032000072256

[6] Jovanović, V. (2016) Trust and Subjective Well-Being: The Case of Serbia. *Personality and Individual Differences*, **98**, 284-288. https://doi.org/10.1016/j.paid.2016.04.061

[7] 袁正, 夏波. 信任与幸福: 基于 WVS 的中国微观数据[J]. 中国经济问题, 2012(6): 65-74.

[8] Kausto, J., Elo, A.L., Lipponen, J., *et al*. (2005) Moderating Effects of Job Insecurity in the Relationships between Procedural Justice and Employee Well-Being: Gender Difference. *European Journal of Work and Organizational Psychology*, **14**, 431-452. https://doi.org/10.1080/13594320500349813

[9] Cassar, V. and Buttigieg, S.C. (2015) Psychological Contract Breach, Organizational Justice and Emotional Well-Being. *Personnel Review*, **44**, 217-235. https://doi.org/10.1108/PR-04-2013-0061

[10] 赵继新, 吴萌萌. 组织公平对员工工作幸福感的影响研究[J]. 北方工业大学学报, 2021, 33(2): 16-25.

[11] 万广华, 张彤进. 机会不平等与中国居民主观幸福感[J]. 世界经济, 2021, 44(5): 203-228.

[12] Sun, F. and Xiao, J.J. (2012) Perceived Social Policy Fairness and Subjective Wellbeing: Evidence from China. *Social Indicators Research*, **107**, 171-186. https://doi.org/10.1007/s11205-011-9834-5

[13] 徐淑一, 陈平. 收入, 社会地位与幸福感——公平感知视角[J]. 管理科学学报, 2017, 20(12): 99-116.

[14] 黄嘉文. 收入不平等对中国居民幸福感的影响及其机制研究[J]. 社会, 2016, 36(2): 123-145.

[15] 王洁菲, 姚树洁. 收入差距、努力指数与居民主观幸福感[J]. 南开经济研究, 2022(4): 3-21.