

# 基于随机森林预测彩礼返还比例

湛靖之

广西师范大学数学与统计学院, 广西 桂林

收稿日期: 2026年6月24日; 录用日期: 2026年7月16日; 发布日期: 2026年7月29日

## 摘要

彩礼是我国民间婚俗的重要组成部分, 近年来因婚约解除引发的返还彩礼纠纷日益受到关注, 但基于真实裁判文书数据的量化研究仍较为缺乏。本文从中国裁判文书网获取湖南省2024年涉及彩礼返还问题的104份民事判决书, 系统提取了双方是否领证、是否举办婚礼、是否同居、共同生活时间、子女生育情况、各项彩礼及相关财物金额等关键事实特征。首先采用K均值聚类分析当事人群体类型, 进而构建随机森林回归模型对彩礼正礼、三金五金及红包礼金的返还比例进行预测与对比。结果表明, 彩礼正礼金额、子女生育个数、共同生活时间是影响返还比例的最重要因素; 而是否领证、是否举办婚礼的影响相对微弱。本文为彩礼返还纠纷的司法裁判提供了数据驱动的参考, 也为当事人理性预估诉讼结果提供了实证依据。

## 关键词

彩礼, K-Means, 随机森林

# Predicting the Return Ratio of Bride Price Based on Random Forest

Jingzhi Chen

School of Mathematics and Statistics, Guangxi Normal University, Guilin Guangxi

Received: June 24, 2026; accepted: July 16, 2026; published: July 29, 2026

## Abstract

Bride price is an important component of traditional Chinese marriage customs. In recent years, disputes over the restitution of bride price arising from the dissolution of engagements have attracted growing public attention, yet quantitative studies based on authentic court judgment data remain scarce. This paper retrieves 104 civil judgments from Hunan Province in 2024 that involve bride price restitution issues from the China Judgments Online website, and systematically extracts

key factual characteristics, including whether the parties obtained a marriage certificate, whether a wedding ceremony was held, whether they cohabited, the duration of cohabitation, the number of children born, and the amounts of various bride-price items and related properties. First, K-means clustering is used to analyze the types of party groups; then a random forest regression model is constructed to predict and compare the restitution ratios of the core bride price, gold jewelry (customarily referred to as the “three golds” and “five golds”), and red-envelope cash gifts. The results indicate that the amount of the core bride price, the number of children, and the cohabitation duration are the most influential factors affecting the restitution ratio, whereas the effects of whether a marriage certificate was obtained and whether a wedding ceremony was held are relatively minor. This study provides a data-driven reference for judicial adjudication in bride-price-restitution disputes and offers empirical evidence for the parties to rationally estimate possible litigation outcomes.

## Keywords

Bride Price, K-Means, Random Forest

Copyright © 2026 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

彩礼，作为中国传统婚俗中的重要组成部分，原本承载着缔结两姓之好、表达诚意与祝福的社会功能。然而，随着社会婚恋观念的变迁，尤其是年轻一代中“不领证先同居”“试婚”等新型交往模式的普及，彩礼纠纷在近年来呈上升趋势。当双方未能最终登记结婚时，已给付的彩礼是否应当返还、返还多少，往往成为对簿公堂的核心争议。这一问题不仅关系到当事人切身利益，也折射出民间习惯与法律规定之间的张力。

学术界对彩礼返还问题的关注由来已久，形成了较为丰富的研究成果。从彩礼的法律性质看，学界的通说将其界定为“附解除条件的赠与”或“以缔结婚姻为目的的赠与”。胡弓卿正(2023)对司法实践中彩礼返还的疑难问题进行系统梳理，指出各地区法官在审理此类案件时，普遍根据当地风俗、过错推定、共同生活时间等因素行使自由裁量权，导致裁判尺度不统一，对判决的公正性和权威性产生影响[1]。胡云红与宋天一(2022)以全国法院 158 份问卷调查和相关裁判文书为对象，对彩礼返还纠纷的法律适用进行了深入实证研究，揭示了司法实践中普遍存在的法律漏洞补充难题[2]。康娜(2025)则以山东省为例，考察了婚约彩礼习惯与制定法之间的冲突与协调机制[3]。

从区域性司法实践的视角来看，部分研究已开始关注特定地区的裁判倾向。有研究以 2019 年湖南省 31 份裁判文书为样本，分析了彩礼返还司法适用规则中存在的问题，指出现有司法解释不够细致、缺乏统一认定标准，导致实务中彩礼纠纷案件难以得到较好解决[4]。本文同样聚焦湖南省，但将样本扩展至 2024 年 104 份判决书，在时间跨度和样本规模上均有所突破。

从法律规范层面看，2024 年最高人民法院发布的《关于审理涉彩礼纠纷案件适用法律若干问题的规定》(法释〔2024〕1 号)具有里程碑意义。该司法解释首次系统明确了彩礼的认定范围、诉讼主体资格、返还比例与方式，并增设了“共同生活”作为重要的考量标准。其规定第六条指出，双方未办理结婚登记但已共同生活的，人民法院应根据彩礼实际使用及嫁妆情况，综合考虑共同生活及孕育情况、双方过错等事实，结合当地习俗，确定是否返还以及返还的具体比例。这一规定为本文的特征选取提供了重要

的法律依据——本文提取的“共同生活时间”“子女生育情况”“是否举办婚礼”等变量，正是该司法解释所强调的核心考量因素。与此同时，学界对上述司法解释的回应也十分及时，已有研究从法律解读与适用分析的视角深入剖析了“共同生活”标准增设的理论与实践意涵[5]。

然而，上述法律规范与司法实践研究主要停留在定性分析和个案评析层面，基于大规模裁判文书数据对彩礼返还比例进行量化预测的研究仍付之阙如。这正是本文试图填补的空白。近年来，机器学习方法在司法判决预测领域取得了显著进展，为该类研究提供了新的范式。法律判决预测(Legal Judgment Prediction, LJP)迅速发展为自然语言处理和机器学习交叉领域的研究热点。Dina 等人(2025)对 LJP 研究进行了系统的文献综述，归纳了罪名预测、法条适用、刑期预测三大任务方向，总结了 TF-IDF 等 21 种自然语言处理方法以及支持向量机(SVM)等主流机器学习方法的应用现状，同时指出数据量不足、分类性能偏低、数据可获取性差、标签标注困难等是该领域面临的主要挑战[6]。国内研究方面，已有学者提出基于多任务学习的民事案件判决预测方法，利用案由与法条之间的依赖关系实现联合判决预测[7]。

本文将同时构建随机森林回归模型，对彩礼正礼、三金五金及红包礼金的返还比例进行预测，揭示各影响因素对返还比例的边际效应方向与强度，帮助当事人和司法实务工作者理性预估类似纠纷的裁判结果，也为广大适婚同胞提供参考。

## 2. 分析过程

本文所使用的数据均来自中国裁判文书网(<https://wenshu.court.gov.cn/>)，通过关键词“返还彩礼”筛选湖南省 2024 年的文书可得共 114 份民事判决书。排除个别含有关键词但与研究无关的案例或重复案例(二审维持原判时去除一审数据)，对判决书中的信息进行提取可得 104 份数据。

在提取数据时，只参考判决书中法院认定事实部分与判决结果部分。本文提取了判决书数几审(取值 1 或 2)，是否维持原判(是取 1，否取 0，一审默认为 0)，上诉人/原告性别(男取 1，女取 0)，一审中是否为原告(是取 1，否取 0，一审默认为 1)，男女双方是否领证(是取 1，否取 0)，男女双方是否举办婚礼(是取 1，否取 0)，男女双方是否同居(是取 1，否取 0)，男女双方共同生活时间(以年为单位，不满一年计 0，满一年四舍五入)，男女双方人流、试管、怀孕未出生、已出生子女个数(例：试管两次、自然怀孕但流产、自然怀孕已出生一名子女、自然怀孕未出生一名子女的情况记为 5)，彩礼正礼数额(单位：元)，彩礼正礼返还比例(%)，三金、五金等大额首饰金额(单位：元)，三金、五金等大额首饰返还比例(%)，见面礼、挂钱等女方及女方亲戚收取红包金额(单位：元)，红包返还比例(%)，男方总支出(单位：元)，女方总返还(单位：元)。该部分的关键事实信息提取参考了文献[1]中的变量选择。

需说明的是，本文主要站在当事人角度考虑问题，在 104 份判决书中，存在男方给予彩礼红包等财物，女方已经部分返还部分彩礼、回礼、已付嫁妆等情况，最极端的一例为双方对簿公堂时女方返回数额已超过法院判定的彩礼数额。故即便法院在部分案例中根据男方累计给予女方的彩礼金额的百分比判定女方应返还数额后再抵扣女方已经返还的数额，本文则在计算彩礼正礼数额等数据时先抵扣女方已返数额，再通过最终判还金额计算返还比。

例如：假设男方支出正礼 50,000 元，女方已回礼 10,000 元，法院判处女方应返还  $50,000 \times 80\% - 10,000 = 30,000$  元。本文计算方式为：彩礼正礼：40,000 元，返还比例： $30,000/40,000\% = 75\%$ 。

同时尤其需要注意本文中见面礼、挂钱等女方及女方亲戚收取红包金额(单位：元)这一项数据。此部分主要包括判决书中经法院经过对证据的确认，认定数额属实的男方支出红包，但需注意其中的一部分往往被认定为与女方亲戚的人情往来，具体表现为非大额(几百到一千)但多量(女方各种叔伯等亲戚有时可达数十人)。这一部分不被法院认定为彩礼但确实属男方支出的部分是本文数据中部分男方总支出会高于判决书认定彩礼总数的主要原因。公平起见，会对同样被事实认定真实数额的女方回礼予以抵扣。

且通过对判决书的阅读对比可发现,若上述给予非直系亲属的红包为大额(多表现为单个在两千元左右及以上),则也会被法院认定为彩礼,但考虑到红包收取方并非女方及父母,难以全部收回,故这一部分几乎不会被判全返。

首先宏观来看上诉人(原告)的性别与该次判决属于几审的关系。对 104 份数据而言,其中一审数据共 81 份,其中原告为男性的占比为  $79/81\% = 97.53\%$ ,这一比例显然符合人们的常识认知。一审中唯二的两例原告为女性的案例,其中一项为男女双方本已私下协商完毕,女方后悔多返,判决结果为不支持女方主张,而另一例为女方起诉离婚时男方要求返还彩礼。

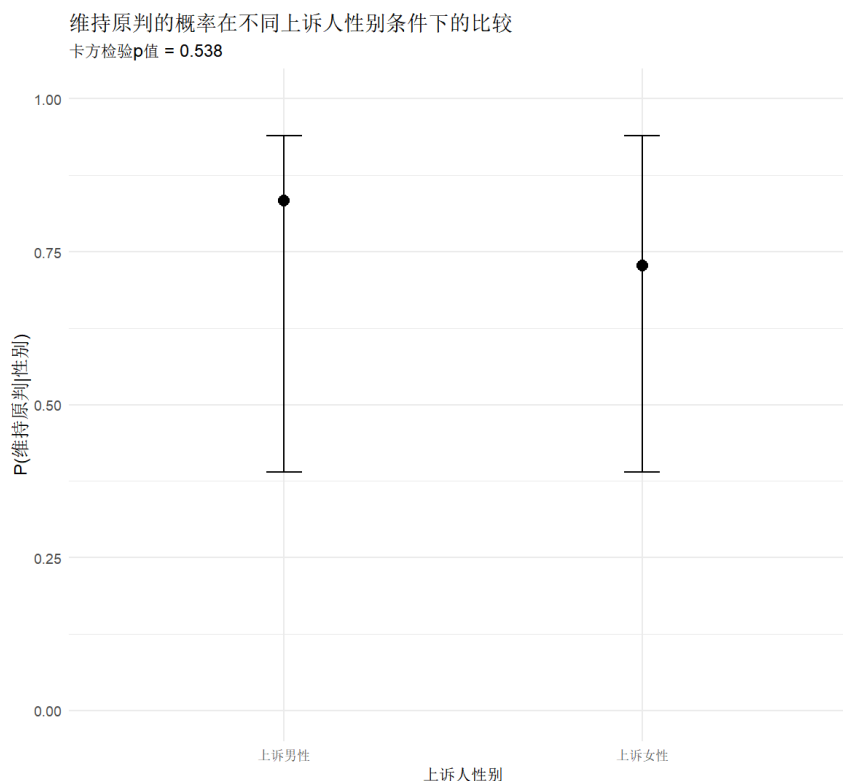
而 23 例二审数据中上诉人的性别比例与一审截然不同。二审中上诉人性别与二审是否维持原判的比例可见表 1。

**Table 1.** The second-instance affirmation rate of the original judgment

**表 1.** 二审是否维持原判的比例

| 性别 | 是否维持原判 | 人数 | 比例     | 总人数 | 总比例    |
|----|--------|----|--------|-----|--------|
| 男  | 是      | 10 | 43.48% | 12  | 52.73% |
|    | 否      | 2  | 8.70%  |     |        |
| 女  | 是      | 8  | 34.78% | 11  | 47.83% |
|    | 否      | 3  | 13.04% |     |        |

对其进行二元数据相关性检验,可得到图 1 中的结果。



**Figure 1.** Comparison of affirmation probability of the original judgment under different gender conditions of the appellants

**图 1.** 维持原判的概率在不同上诉人性别条件下的比较

可以看到，虽然图中显示当上诉人性别为男性时维持原判的概率要高于女性，但二者的置信区间发生明显重叠，且卡方检验的  $p$  值远大于 0.05，故认为该样本中没有足够的证据证明上诉人的性别与二审是否维持原判之间存在明显关联。可认为是由于样本量较小造成的后续条件概率估计存在较大差距，而置信区间基本重叠的情况。

而基于常规逻辑以及判决书中上诉人的诉求，可以认为上诉人申请二审即意味着认为一审判决结果不利于己方利益，而二审维持原判则表明二审法院不支持上诉人的诉求。故可认为在 2024 年湖南省上传裁判文书网的二审判决中不存在法院在二审中更偏向女性上诉人的情况。

现根据男女双方相处产生的客观事实，即男女双方是否领证(是取 1，否取 0)，男女双方是否举办婚礼(是取 1，否取 0)，男女双方是否同居(是取 1，否取 0)，男女双方共同生活时间(以年为单位，不满一年记 0，满一年四舍五入)，男女双方人流、试管、怀孕未出生、已出生子女个数五个变量进行聚类分析可得图 2 以及表 2 中的结果。

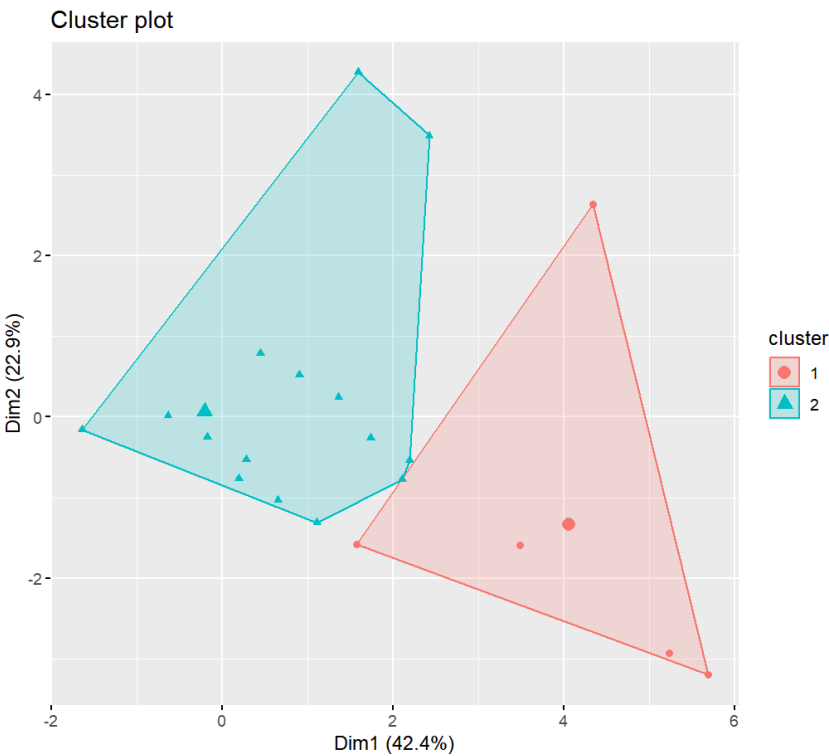


Figure 2. Cluster analysis  
图 2. 聚类分析

Table 2. Cluster centers  
表 2. 聚类中心点

| 变量                      | 聚类 1 中心点 | 聚类 2 中心点 |
|-------------------------|----------|----------|
| 男女双方是否领证                | 0.2      | 0.04     |
| 男女双方是否举办婚礼              | 0.8      | 0.26     |
| 男女双方是否同居                | 1.0      | 0.76     |
| 男女双方共同生活时间              | 4.6      | 0.43     |
| 男女双方人流、试管、怀孕未出生、已出生子女个数 | 1.8      | 0.17     |

K-means 聚类算法通过计算每个聚类在变量上的均值(中心值), 揭示不同用户群体的核心特征。这些中心值反映了各聚类内样本在特定变量上的平均水平或分布比例, 根据以上结果我们可以总结两类当事人的特点。第一类为同居且共同生活较长时间, 交往期间女方怀过孕过且多数已按照民间习俗举办婚礼。第二类为多数同居过, 但共同生活时间较短, 女方多数未怀过孕。这两种情况也符合直观逻辑与认知。

接下来, 我们想根据上述聚类中的变量以及实际案例中男方给予女方的各项与彩礼相关的财物金额来模拟预测最后法院判还的比例。为了方便计算, 在处理数据时对变量进行表 3 中的重新命名。

Table 3. Variable renaming  
表 3. 变量重命名

| 重命名 | 原名                           | 取值        |
|-----|------------------------------|-----------|
| lz  | 男女双方是否领证(是取 1, 否取 0)         | 0 或 1     |
| hl  | 男女双方是否举办婚礼(是取 1, 否取 0)       | 0 或 1     |
| tj  | 男女双方是否同居(是取 1, 否取 0)         | 0 或 1     |
| sj  | 男女双方共同生活时间(以年为单位, 不满一年记 0)   | 0 或正整数    |
| zv  | 男女双方人流、试管、怀孕未出生、已出生子女个数      | 0 或正整数    |
| X1  | 彩礼正礼数额(单位: 万元)               | 实数        |
| X2  | 三金、五金等大额首饰金额(单位: 万元)         | 实数        |
| X3  | 见面礼、挂钱等女方及女方亲戚收取红包金额(单位: 万元) | 实数        |
| a   | 彩礼正礼数额返还比例                   | 大于 0 小于 1 |
| b   | 三金、五金等大额首饰金额返还比例             | 大于 0 小于 1 |
| c   | 见面礼、挂钱等女方及女方亲戚收取红包金额返还比例     | 大于 0 小于 1 |

• 本文的目的是使用上述除 a、b、c 外的变量来预测 a、b、c 的值。由于本文采用的湖南省 2024 年上传裁判文书网的样本共 104 组, 样本量较小并不能很好地支撑如神经网络等模型方法。于是针对 a、b、c 分别构建随机森林模型来预测。图 3 展示了影响彩礼正礼返还比例(a)、三金五金返还比例(b)以及红包礼金返还比例(c)的各特征变量的重要性程度(以均方误差 MSE 增加量为衡量标准)。图中横轴为变量重要性得分, 纵轴为各输入变量, 从图 3 中可以看出:

1. 对于彩礼正礼返还比例(a), 最重要的三个变量依次为: 子女生育个数(zv)、彩礼正礼金额(X1)、共同生活时间(sj)。这表明, 女方是否怀孕或生育子女, 以及共同生活的实际时长, 是法院判决正礼返还比例时最关键的考量因素。

2. 对于三金五金返还比例(b), 三金金额(X2)的重要性显著高于其他变量, 说明此类财物的返还比例几乎完全由其自身金额决定, 受其他因素(如是否领证、是否办婚礼)影响较小。

3. 对于红包礼金返还比例(c), 共同生活时间(sj)和是否同居(tj)的重要性较高, 而子女生育个数(zv)的影响相对较弱。这反映出法院在处理与亲属人情往来相关的红包返还时, 更关注双方实际共同生活的状态, 而非生育情况。

此外, 是否领证(lz)和是否举办婚礼(hl)在所有三个目标变量中重要性排名均靠后, 说明在湖南省 2024 年的裁判实践中, 形式上的婚姻登记与婚礼仪式对返还比例的影响远小于实质性共同生活与生育情况。

图 4 展示了随机森林模型在测试集上对 a、b、c 三个返还比例的预测结果与真实判决结果之间的对比, 理想情况下, 若所有点应紧密分布在  $y=x$  对角线附近, 则表示预测值与真实值完全一致。彩礼正礼返还比例(a)的对比图中散点整体沿对角线分布, 尤其在真实值处于 50%至 80%区间时, 预测值与真实值较为接近。但在真实值接近 0%或 100%的极端区域, 部分点偏离对角线较远, 说明模型对全返或不返的

极端判决预测能力有限。而三金五金返还比例(b)以及红包礼金返还比例(c)的对比图中散点分布较彩礼正礼返还比例(a)更为分散,多数点的真实值取 0 或 1 导致了散点大量远离对角线,说明模型未能正确识别极端情况。这表明模型对于小额财物(红包)以及大额首饰(三金五金)的返还比例预测效果不理想,主要可能是这两类财物的判决受个案中具体人情往来、举证情况等难以量化的因素影响较大,且判决结果中全部返还和全部不返还的情况占多数。

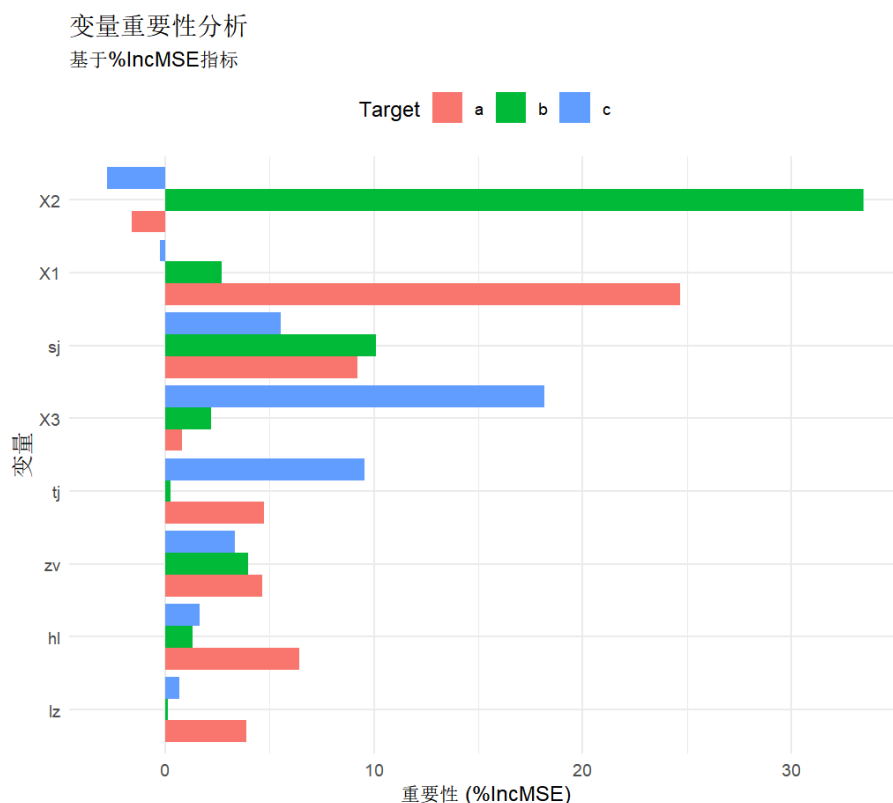


Figure 3. Variable importance analysis

图 3. 变量重要性分析

随机森林模型下真实值与预测值对比

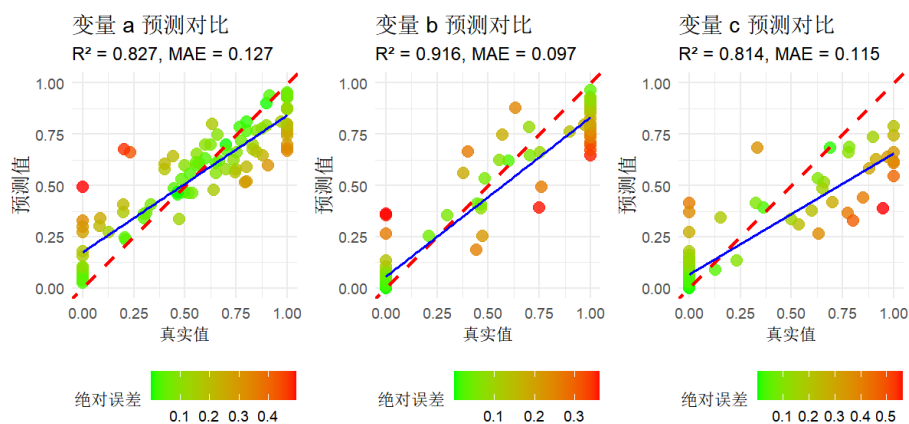


Figure 4. Random forest prediction results—comparison of predicted values and actual values

图 4. 随机森林预测结果——预测值与真实值对比

综合来看,当前模型仅对彩礼正礼返还比例(a)具有中等程度的预测参考价值,对极端值及非正礼类财物的预测效果较差。

### 3. 结论

从预测结果与真实值对比图可看出,该模型仍存在较大的改进空间。而导致这一现象的原因可能出在两方面。首先对裁判文书网对非法律从业者只开放目录中的前 600 项数据,加上其政府网站的性质导致难以通过爬虫快速收集信息与数据,这样的现状导致本文获取数据十分困难,最终只选取了湖南省 2024 年的数据进行分析。如果后续想要改进,可以优先从扩大研究的地点和时间跨度出发增加样本数据量。

而基于目前的结果可见到彩礼正礼金额(X1),三金、五金等大额首饰金额(X2),见面礼、挂钱等女方及女方亲戚收取红包金额(X3)分别对彩礼正礼返还比例(a)、三金五金返还比例(b)、红包礼金返还比例(c)十分重要,即本文中区分的三类“彩礼”的金额是会相互影响返还比例的。而是否已拥有子女、共同生活时间的长短以及是否同居相对来说会更影响彩礼返还的比例。其中,子女的有无及个数对正礼的返还比例影响最大,共同生活时间与是否同居更影响见面礼、红包礼金等的返还比例,三金、五金等大额首饰的返还比例几乎只与其金额相关,受到其他变量的影响较小。而是否举办婚礼、是否领证登记结婚对最终结果的影响并不算大。

由于本文在进行分析时只选取了部分关键信息转化为变量而忽略了判决书中大量描述性文字,因此出现了部分与直觉或传统认知相悖的结果,想要合理地解释结论就需要回到判决书本身。

首先,是否领证登记结婚对返还彩礼的比例影响不大,这个结论无疑看起来违反常识。但如果回到判决书本身就能发现,在这 104 例判决书中,当事人领证结婚且举办婚礼的仅有 5 例,而这 5 例中仅有 1 例当事人共同生活时间达到了 6 年,系女方起诉离婚。其余 4 例案件中当事人的共同生活时间均未达到 1 年。这说明有关返还彩礼的案件中,当事人真正领证的情况很少,且几乎都是新婚夫妻闹矛盾,既然案件数量占比小,对判决结果的影响自然不大。事实上,如果当事人结婚时间较长,婚后财产问题复杂,那么当事人更可能打离婚官司而非返还彩礼问题的官司。

其次,是否举办婚礼对返还彩礼比例的影响同样不大,这一结论同样可能违背了部分人的常识。但首先值得注意的是,我国的民法典第 1049 条明确规定了结婚登记是确立婚姻关系的唯一途径,婚礼只是民间习俗,不具备法律效力。认为举办了婚礼的二人已经确认了夫妻关系,这是一种民间常见的谬误,因此是否举办婚礼本就不可能在返还彩礼问题上起到决定性作用。而在湖南省 2024 年的这 104 例案件中,包括领证结婚的 5 例在内,共有 30 例案件中当事人举办了婚礼,这 30 例中共有 16 例当事人的共同生活时间至少达到了 1 年,超过 1 年的则只有 8 例,可见举办婚礼也无法说明当事人已经共同生活了较长的时间。同时,在这举办婚礼的 30 例案件中,还存在女方订婚时仅 15 岁、女方订婚及生育时均未到法定婚龄、男方订婚时未到法定婚龄、当事人双方已和平解决而女方父母再上诉、女方不孕不育等多例特殊案件。显然在上述几种特殊案件中,举办婚礼与否不可能是最核心的问题。

三金、五金等大额首饰的返还比例几乎只与其金额相关,受到其他变量的影响较小,这一点也需要结合判决书中的具体情况讨论。本文讨论的 104 例案件来源于湖南省地不同市县区域,必须明确不同的法院往往采取不同的判决习惯。在本文讨论的案件中,对三金、五金等大额首饰的判决方法可以大致分为 3 种。第一种为返还实物或按照购买价格返还金钱(即百分百返还);第二种为对三金、五金等大额首饰计价后单独设置返还的比例(例如男方送出的三金、五金等大额首饰合计价值 1 万元,判决女方返还 7 千元);第三种为将三金、五金等大额首饰计价后与其他彩礼款项相加,只判决对总金额的返还比例(例如男方给出正礼 8 万元,大额首饰 2 万元,共计 10 万元,判决女方返还 $(8 + 2) \times 80\% = 8$  万元)(上述例子中的数据仅供理解,不来源于判决书案例)。因此,针对第三种情况是无法直接得到三金、五金等大额首饰

的具体归还比例的, 本文在分析时采取的是直接以总体返还比例来代替, 本就会存在误差。除此之外, 从这 104 例判决书中找不到首饰能被算作三金、五金等大额首饰的具体标准, 因此在案例中, 三金、五金等大额首饰的返还情况具有特别强的地域差异, 本文中的有关结论仅适合作为参考。

从以上结果不难看出, 如果男方希望自己以结婚为前提给予的彩礼在未达成理想结果情况下可以尽可能地收回, 那么首先就应当控制彩礼的金额。其次尽量避免在领证结婚前长时间与女朋友共同生活, 尤其是要避免婚前怀孕。而女方也不应抱着敛财的目的去收取高额彩礼, 收取彩礼最后没结成婚也应主动返还应当返还的部分, 而不是无限度地把较大金额的财物, 例如上万元的金器首饰, 说成是男方无条件赠予。

## 参考文献

- [1] 胡弓卿正. 司法实践中彩礼返还的疑难问题探析[J]. 江西警察学院学报, 2023(6): 122-128.
- [2] 胡云红, 宋天一. 彩礼返还纠纷法律适用研究——以全国法院 158 份问卷调查和相关裁判文书为对象[J]. 中国政法大学学报, 2022(6): 5-27.
- [3] 康琳. 彩礼纠纷司法审判的功能预期、现实困境与脱困路径[J]. 江西警察学院学报, 2025(4): 78-84.
- [4] 孟丽媛. 彩礼返还司法适用规则研究——以 2019 年湖南省 31 份裁判文书为样本[J]. 现代商贸工业, 2021, 42(32): 112-113.
- [5] 李燕, 沙媛. 彩礼返还最新规定的法律解读与适用分析[J]. 西部学刊, 2025(1): 63-66.
- [6] Dina, N.Z., Ravana, S.D. and Idris, N. (2025) Legal Judgment Prediction Using Natural Language Processing and Machine Learning Methods: A Systematic Literature Review. *Sage Open*, **15**, 1-25.  
<https://doi.org/10.1177/21582440251329663>
- [7] 余连玮, 马志柔, 刘杰, 等. 基于多任务学习的民事案件判决预测方法[J]. 计算机应用与软件, 2024, 41(1): 18-25.