

关于数据安全关注度调查研究 ——以电子取证为例

胡闪闪¹, 李玉莲^{2*}, 陈心雨², 孙天琦², 刘译丹¹

¹塔里木大学网络安全学院, 新疆 阿拉尔

²塔里木大学信息工程学院, 新疆 阿拉尔

收稿日期: 2025年7月19日; 录用日期: 2025年8月9日; 发布日期: 2025年8月21日

摘 要

信息技术发展使数据安全成为社会焦点, 数据泄露威胁个人隐私、商业机密和国家安全。本研究通过问卷调查分析公众对电子取证数据安全的认知, 发现呈现“基础较好但实践不足”特点: 72.39%受访者知晓概念, 仅28.76%认为公众关注度高。学历显著影响认知水平($\chi^2 = 23.030, p < 0.001$), 硕士以上群体风险识别准确率(89%)高于高中以下群体(58%)。数据安全重视程度与关注度正相关($p < 0.05$)。决策树和有序Logistic模型分析显示, 信任度、数据安全重要性认知及年龄显著影响关注度。研究为数据安全政策制定等提供依据, 创新性结合关注度调查与电子取证技术, 为构建防护体系提供新思路。

关键词

数据安全, 关注度调查, 电子取证, 数据安全技术

Investigation and Research on the Attention Paid to Data Security

—Taking Electronic Forensics as an Example

Shanshan Hu¹, Yulian Li^{2*}, Xinyu Chen², Tianqi Sun², Yidan Liu¹

¹College of Cyber Security, Tarim University, Alaer Xinjiang

²College of Information Engineering, Tarim University, Alaer Xinjiang

Received: Jul. 19th, 2025; accepted: Aug. 9th, 2025; published: Aug. 21st, 2025

Abstract

The development of information technology has made data security a social focus. Data leakage

*通讯作者。

文章引用: 胡闪闪, 李玉莲, 陈心雨, 孙天琦, 刘译丹. 关于数据安全关注度调查研究[J]. 统计学与应用, 2025, 14(8): 263-271. DOI: 10.12677/sa.2025.148233

threatens personal privacy, business secrets and national security. This study analyzed the public's awareness of the security of electronic forensic data through a questionnaire survey and found that it presented the characteristics of "good foundation but insufficient practice": 72.39% of the respondents were aware of the concept, while only 28.76% believed that the public attention was high. Educational attainment significantly affects cognitive level ($\chi^2 = 23.030$, $p < 0.001$). The accuracy rate of risk identification in the group with a master's degree or above (89%) is higher than that in the group below high school (58%). The degree of emphasis on data security is positively correlated with the degree of attention ($p < 0.05$). Decision tree and ordered Logistic model analysis show that trust, perception of the importance of data security, and age significantly affect attention. The research provides a basis for the formulation of data security policies and innovatively combines attention surveys with electronic forensics technology, offering new ideas for building a protection system.

Keywords

Data Security, Attention Survey, Electronic Forensics, Data Security Technology

Copyright © 2025 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. 研究背景与意义

在数字技术迅猛发展、数据爆炸式增长的当下，大数据运用引发诸多个人隐私与信息安全问题[1]。网络攻击、数据泄露频发，我国虽出台相关法规加大技术投入，但数据特性使其防护艰难，新技术又加剧安全风险的复杂隐蔽性。开展数据安全关注度调查研究意义非凡。数字经济下数据是关键要素，开展数据安全关注度调查，可助政府施策、国安机关防范窃密。数据作为生产要素受到各方的高度关注，企业面临的数据安全态势也在不断地发生着变化[2]。准确把握社会各界对数据安全的关注度和需求，有助于政府部门在利用数据提升治理效能的同时，加强对数据的安全管理和保护。例如，在智慧城市建设中，涉及大量居民的个人信息和城市运行数据[3]。综上所述，此次调研对推动数字经济发展、维护国家安全与社会稳定具有积极意义。

2. 研究方法

2.1. 调查问卷

本文采用问卷调查法。此次数据安全关注度调查研究有四个要点：一是了解被调查者对电子取证中数据安全概念，如信息泄露、数据修改的认知情况。二是观察其最关注数据收集、保存、使用哪个阶段，以及对合法性、保密性、完整性的重视倾向。三是分析信息泄露经历、工作接触数据频繁度对其重视程度的影响。四是调查面对数据安全问题的反应及收集保障建议。调查对象覆盖多类人群，通过广泛调研这些不同年龄、职业、教育背景的群体，力求精准反映社会各阶层对数据安全的关注情况，为后续研究筑牢数据根基。

2.2. 决策树模型分析

本次研究采用决策树中的 CART (Classification and Regression Trees) 生长法构建模型，从根结点开始，对实例的某一特征进行测试，根据测试结果，将实例分配到其子结点；这时，每一个子结点对应着该特

征的一个取值。其中,调查问卷中“您是否信任当前电子取证技术的安全性”这一问题作为核心特征优先进入根节点测试,这里的“信任”指公众对电子取证技术在数据保护、防篡改、隐私性等方面的可靠感知,包含对技术稳定性、执行规范性及结果公正性的综合认可。如此递归地对实例进行测试并分配,直至达到叶结点。最后将实例分到叶结点的类中[4]。这种算法既可以用于分类,也可以用于回归问题。以“您认为当前公众对电子取证数据安全的关注度如何”作为因变量,将性别、年龄、职业、学历、是否听说过电子取证数据安全概念以及是否信任当前电子取证技术的安全性作为自变量。在模型构建过程中,设定最大树深度为 5,父节点中的最小个案数为 10,子节点中的最小个案数为 5,以此控制模型的复杂度,避免过拟合现象的发生。同时,运用交叉验证对模型进行验证,确保模型的稳定性和可靠性。

2.3. 有序 Logistic 模型分析

本研究采用 Logistic 连接函数构建有序 Logistic 模型。该模型基于累计概率的对数变换,能够有效处理因变量为有序分类变量的情况,根据所得到的数据,参考文献和相关规定,选取合适的自变量,建立样本和自变量之间的回归模型,再通过模型对未知的样本数据进行分析和预测[5]。

3. 问卷分析

3.1. 信度, 效度检验

本研究共收集 520 份问卷,有效问卷 518 份,有效率 99.6%。性别较均衡,女性略多,占 51.4%,男性占 48.6%。年龄分布以青年和中年为主,46 岁及以上仅占 14.9%。职业方面,企业员工(非 IT)最多,占 48.1%,学生次之(31.1%),IT/网络安全从业者仅 14.1%。学历在本科及以上学历者占 66.6%,显示高学历群体对数据安全认知较全面。详情见图 1。

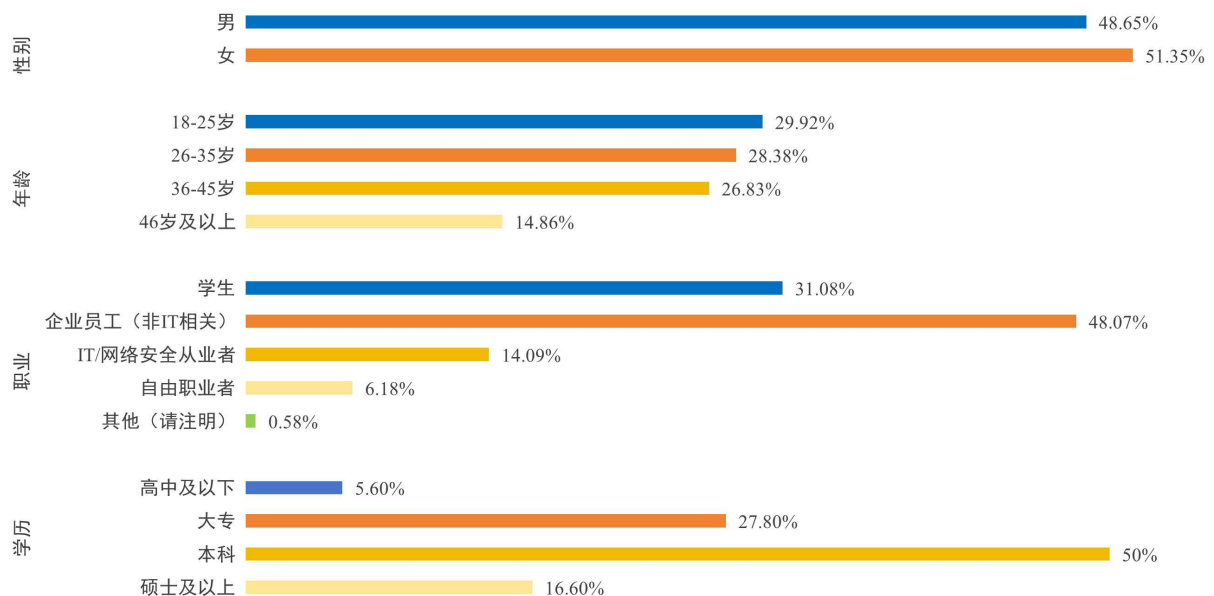


Figure 1. Distribution map of basic information of the survey sample

图 1. 调查样本基本信息分布图

从表 1 可以得出,该测量指标的 Cronbach's α 信度系数为 0.913,按照信度指标判别标准,高于 0.9,说明各指标的可靠性符合条件,具有较好的内部一致性。效度分析显示, KMO 值为 0.927,并且 Bartlett

球度检验结论给出的 p 值近似为 0。调查问卷中的变量之间的相关性较强，数据呈现出高度的相关性和有效性。

Table 1. Reliability and validity test table
表 1. 信度，效度检验表

指标		数值
Cronbach's α 系数		0.913
KMO 值		0.927
Bartlett 球形度检验	近似卡方	2337.436
	df	21
	p 值	0

注：样本量 = 518，项数 = 7。

3.2. 数据安全认知统计分析

根据统计，只有 28.76% 的受访者认为公众对电子取证数据安全的关注度较高或非常高，关注度较低的比例达到 16.79%。整体来看，公众对电子取证数据安全的关注度普遍偏低。具体数据如图 2 所示。

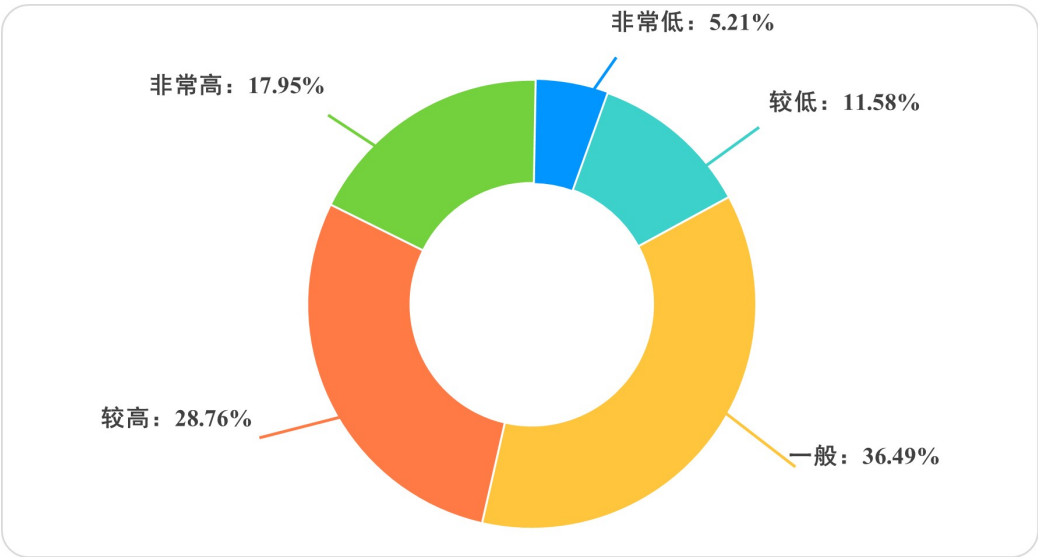


Figure 2. The survey sample indicates the current public concern over the security of electronic forensic data
图 2. 调查样本中认为当前公众对电子取证数据安全的关注度

利用卡方检验研究职业，认为电子取证过程中数据安全重要性对于数据安全关注度的差异关系，从表 2 可看出：不同性别、年龄、职业和学历对于数据安全关注度不会表现出差异性($p > 0.05$)，意味着不同性别、年龄、职业和学历对于数据安全关注度均表现出一致性，并没有差异性。不同观点对于电子取证数据安全关注度表现出明显差异性($p < 0.05$)，说明越认为过程重要的人对电子取证数据安全的关注度越高。

3.3. 电子取证统计分析

如图 3 所示，在调查中有 72.39% 的受访者表示他们听说过电子取证数据安全的概念，仅有 27.61% 的

受访者表示未听说过。这表明该概念在受访者中具有较高的认知度。在对信任度调查中,选择“一般信任”“信任”和“非常信任”的受访者合计占比达 90.92%。这表明大多数受访者对电子取证技术持积极态度。

Table 2. Cross-table chi-square analysis table of attention, basic information and importance

表 2. 关注度与基本信息及重要性交叉表卡方分析表

题目	χ^2	p
性别	8.391	0.078
年龄	15.087	0.237
职业	13.639	0.626
学历	13.865	0.309
您认为电子取证过程中数据安全的重要性如何?	178.547	0.000***

***p < 0.001。

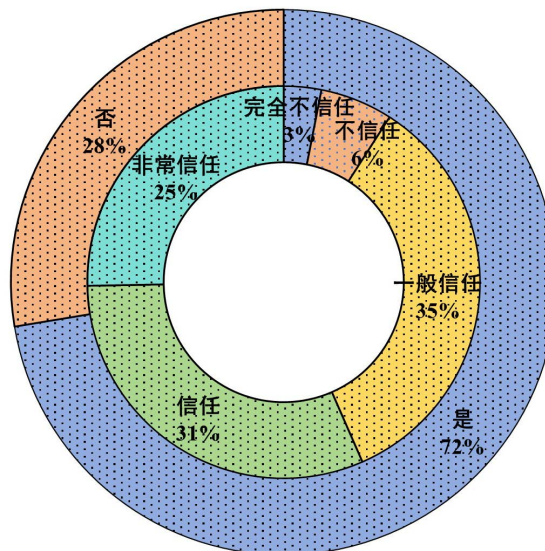


Figure 3. The survey sample indicates the current public awareness and trust in the security of electronic forensic data

图 3. 调查样本中认为当前公众对电子取证数据安全的认知程度及信任度

由表 3 可知,是否听说过电子取证数据安全这个概念对于性别共 1 项不会表现出显著性($p > 0.05$)。另外是否听说过电子取证数据安全这个概念对于年龄,职业,学历是共 3 项呈现出显著性($p < 0.05$)。认为电子取证过程中数据安全的重要性程度不同对于是否信任对当前电子取证技术的安全性信任全部均呈现出显著性差异($p < 0.05$)。

4. 实证分析

4.1. 基于决策树模型的实证分析

本次研究采用决策树中的 CRT 生长法构建模型。决策树以直观的树形结构呈现决策逻辑,无需复杂

Table 3. The chi-square analysis table of the cross-table of cognitive level and basic information, importance and trust
表 3. 认知程度与基本信息，重要性与信任度交叉表卡方分析表

因变量	题目	χ^2	p
是否听说过电子取证数据安全	性别	3.54	0.06
	年龄	10.326	0.016*
	职业	9.951	0.041*
	学历	23.03	0.000***
认为电子取证过程中数据安全的重要性	您是否信任当前电子取证技术的安全性？	350.389	0.000***

*p < 0.05, ***p < 0.001。

的数据预处理，对连续型和分类变量都有良好兼容性，便于理解与应用。该模型旨在从复杂数据中挖掘关键影响因素，通过递归二分优化节点纯度，实现高效的分类或回归预测，同时利用其天然的可解释性，为研究结论提供清晰的规则依据，详情见表 4。

Table 4. Model diagram constructed by CRT growth method
表 4. CRT 生长法构建模型图

指定项		结果	
生长法	CRT		
因变量	您认为当前公众对电子取证数据安全的关注度如何？		您是否信任当前电子取证技术的安全性？您的学历是，您的年龄，您是否听说过电子取证数据安全这个概念？您的职业，您的性别
自变量	您的性别，您的年龄，您的职业，您的学历是，您是否听说过电子取证数据安全这个概念？您是否信任当前电子取证技术的安性？	包括的自变量	
验证	交叉验证		
最大树深度	5	节点数	31
父节点中的最小个案数	10	终端节点数	16
子节点中的最小个案数	5	深度	5

4.2. 决策树模型结果分析

决策树算法是一种基于树形结构进行决策的方法，它通过对数据特征的不断划分，将数据集逐步细分，最终实现对数据的分类或预测[6]。如图 4 所示，本研究通过决策树模型分析发现，公众对电子取证技术安全性的信任度、学历、年龄、是否听说过电子取证数据安全概念、职业和性别是影响公众对电子取证数据安全关注度的重要因素。其中，信任度对关注度的影响最为显著，学历、年龄和信息知晓度也在不同程度上影响着公众的关注度判断。

如表 5 所示，通过交叉验证和重新代入法对模型的风险进行评估，预测公众对电子取证数据安全的关注度一般，正确百分比为 72.5%，模型对于结果预测效果较好。但整体而言，模型在预测公众对电子取证数据安全关注度方面存在一定的误差，预测能力有待进一步提高。

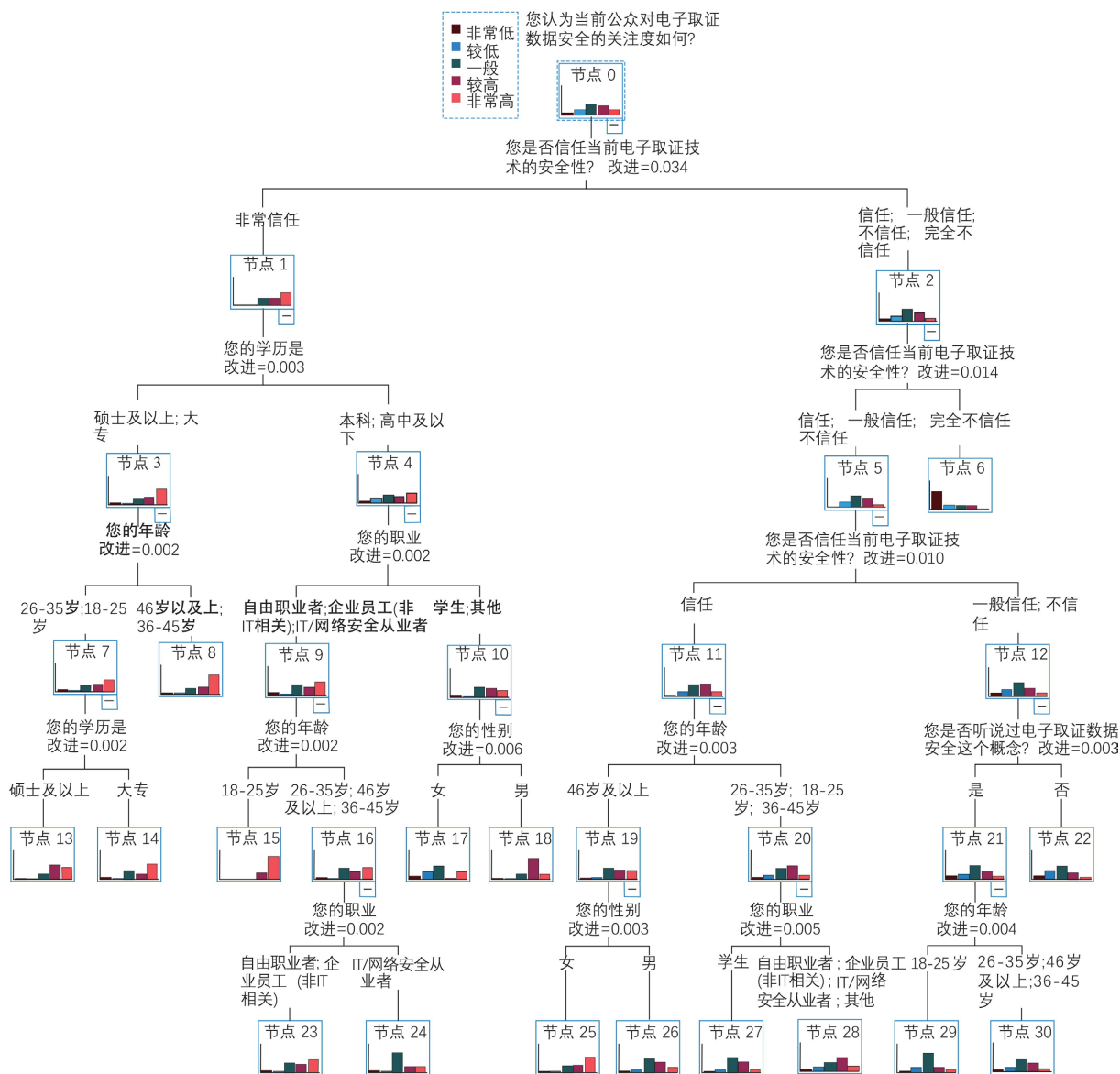


Figure 4. Decision tree model diagram
图 4. 决策树模型图

Table 5. Model prediction capability evaluation form
表 5. 模型预测能力评估表

实测	预测					正确百分比
	非常低	较低	一般	较高	非常高	
非常低	10	0	15	0	2	37.0%
较低	2	0	46	12	0	0.0%
一般	2	0	137	27	23	72.5%
较高	2	0	72	54	21	36.2%
非常高	0	0	21	18	54	58.1%

4.3. 基于有序 Logistic 模型

在构建逻辑回归模型时，首先对选取的自变量进行相关性分析，剔除相关性过高的变量，以避免多重共线性问题。然后，运用最大似然估计法对模型参数进行估计。在评估模型性能时，采用准确率、召回率等指标，综合判断模型对不同信用风险等级的预测能力[7]。根据表 6，似然比检验结果显示，原假设被拒绝($\chi^2 = 152.730$)，($p = 0.000 < 0.05$)。这表明放入的 7 个自变量整体上对因变量具有显著影响，模型构建具有统计学意义，即这些自变量能够有效地解释公众对电子取证数据安全关注度的差异。

Table 6. Likelihood ratio test table for ordered Logistic regression model
表 6. 有序 Logistic 回归模型似然比检验表

模型	-2 倍对数似然值	卡方值	df	p	AIC 值	BIC 值
仅截距	1490.058					
最终模型	1337.327	152.730	7	0.000	1359.327	1406.077

根据表 7，研究显示，公众对电子取证数据安全的关注度受到多种因素的显著影响。在 1%的显著水平下，对电子取证技术安全性的信任、概念认识以及对电子取证过程中数据安全重要性的认知是影响公众关注度的重要因素。信任电子取证技术安全性的公众，其对数据安全的关注度更高；认为数据安全重要性程度越高的公众，对电子取证数据安全的关注度也越高。在 1%显著水平下，“信任技术安全性”“听说过概念”“认为数据安全重要”显著影响关注度，以“是否信任当前电子取证技术的安全性”为例，其 OR 值为 2.422，即信任该技术安全性的公众，对数据安全更关注的概率，是不信任者的 2.422 倍，直观体现该因素对关注度的推动作用，其他显著变量 $OR > 1$ ，说明对应特征也会提升公众关注度，且 OR 越大，影响越强。

$$\ln Y_i = 0.101x_1 + 0.132x_2 + \cdots + 0.261x_7$$

Table 7. Summary table of analysis results of the ordered Logistic regression model
表 7. 有序 Logistic 回归模型分析结果汇总表

项	项	回归系数	标准误	Wald χ^2	p 值	OR 值	OR 值 95% CI
因变量阈值	1.0 (低)	0.646	0.673	0.922	0.337	0.524	0.140~1.960
	2.0 (一般)	2.164	0.670	10.443	0.001	0.115	0.031~0.427
	3.0 (较高)	4.236	0.687	38.045	0.000	0.014	0.004~0.056
	4.0 (非常高)	5.893	0.706	69.597	0.000	0.003	0.001~0.011
自变量	您的性别	0.101	0.164	0.380	0.538	1.107	0.802~1.527
	您的年龄	0.132	0.083	2.550	0.110	1.141	0.970~1.341
	您的职业	0.082	0.099	0.692	0.406	1.086	0.894~1.318
	您的学历是	-0.012	0.103	0.015	0.904	0.988	0.807~1.209
	您是否听说过电子取证数据安全这个概念？	-0.647	0.192	11.420	0.001	0.523	0.360~0.762
	您是否信任当前电子取证技术的安全性？	0.885	0.096	84.275	0.000	2.422	2.005~2.926
	您认为电子取证过程中数据安全的重要性如何？	0.261	0.090	8.349	0.004	1.298	1.088~1.550

5. 结论与建议

本研究基于 518 份有效问卷,运用多种分析方法探讨公众对电子取证数据安全的认知、影响因素及提升路径。发现公众呈现“认知较好但实践不足”特征,群体间有差异,技术信任与认知影响安全行为。认知上,超七成听说过相关概念,但认为关注度高的不足三成,学历、职业影响认知水平。安全行为方面,基础防护普遍,高级措施使用少,46 岁以上防护较差。决策树, Logistic 模型进一步验证多因素对关注度有重要影响,信任度影响尤为突出。

根据调查与分析提出以下建议:围绕提升电子取证数据安全水平,可多维度发力。针对 46 岁以上人群,制作并推送基于真实案例的短视频,传播简单实用的操作方法,以强化其基础防护能力。为高学历群体开设结合电子取证实际案例的演练工作坊,切实提升实操能力。对于企业员工,根据其职业特点和数据接触频率,针对性制定详细的数据安全操作指南,提供具体操作指导。同时,通过开展技术安全评估并公开结果,完善技术制度与法规,开发可视化取证工具,建立安全认证制度,系统性提升技术可信度和公众对数据安全重要性的认知。此外,强化政府、企业、学校协同的安全意识培训,结合不同群体的特征,制定差异化的培训形式与内容,进而整体提升公众对数据安全的关注度与实践水平。

基金项目

塔里木大学校长基金:基于区块链的电子证据固定方法研究(编号:TDZKSS202439);3E 系统的非对称性因果传导效应研究(编号:TDZKSS202417)。

参考文献

- [1] 白娟. 数据安全治理在智慧校园建设中的实践探索[J]. 网络安全技术与应用, 2025(3): 95-97.
- [2] 余强明, 朱小栋, 吴靓, 等. 大数据时代顾客隐私关注度对企业精准营销效益影响的实证研究[J]. 重庆工商大学学报(自然科学版), 2020, 37(4): 95-103.
- [3] 王朋群, 李虎, 欧阳熹, 等. “面向未来有效保护”的数据安全治理[J]. 中国信息安全, 2019(12): 76.
- [4] 李航. 统计学习方法[M]. 北京: 清华大学出版社, 2019.
- [5] 陈晓军, 王启明. 基于有序 Logistic 回归模型的通风系统评价[J]. 煤矿机械, 2024, 45(10): 192-195.
- [6] 凌昊. 基于决策树分类算法的网络入侵检测系统的研究[D]: [硕士学位论文]. 长沙: 湖南大学, 2008.
- [7] 王丽华, 彭定涛. 基于稀疏逻辑回归的信用风险评估模型[J]. 电子商务评论, 2025, 14(1): 1354-1360.
<https://doi.org/10.12677/ecl.2025.141168>