

基于大数据的高职院校贫困生精准认定探析

袁璟瑾¹, 孔淑佩², 陈 燕²

(1. 江苏航运职业技术学院 学生工作处, 江苏 南通 226010;

2. 江苏航运职业技术学院 运输管理与经济学院, 江苏 南通 226010)

摘 要:依托大数据技术实现高职院校贫困生的精准资助是未来学生资助管理改革的趋势。目前, 高职院校精准资助在数据收集、认定工作、跟踪管理等方面还有不少问题, 究其成因主要为评定标准单一, 标签效应负面; 数据整合乏力, 动态管理缺失; 数据统计滥用, 个体考评不足。尝试构建基于大数据的贫困生精准认定工作框架, 推进资助对象认定精准化、资助行为差异化、资助考核合理化, 最后建议采用 SMOTE 方法对资助对象精准认定, 切实提升精准资助工作的成效。

关键词:大数据; 贫困生; 精准认定

中图分类号:G715

文献标志码:A

文章编号:1671-9891(2020)2-0083-05

0 引言

十三五”期间, 习近平总书记指出, 扶贫开发贵在精准, 重在精准, 成败之举在于精准。^[1]贫困生资助是我国扶贫工作的重要组成部分, 随着高职教育的普及, 高职在校贫困生数量也在逐年递增, 不仅加大了贫困生资助的工作量, 也增加了精准资助难度系数。目前, 传统的资助模式已经不能完全满足高职院校的精准资助要求。在新时代科技蓬勃发展的背景下, 如何利用信息化手段采集学生基础数据信息, 利用数据分析技术助力高校贫困生精准认定, 保障高职院校贫困生顺利完成学业, 已成为一个值得深入探讨的问题。

1 高职院校贫困生精准认定工作面临的困境

1.1 数据收集不客观, 资助缺乏精度

全面采集贫困生数据, 找准资助对象, 是高职院校贫困生资助的起点, 也是难点。目前, 高职院校贫困生认定主要参考教育部、财政部《关于进一步加强和规范高校家庭经济困难学生认定工作的通知》《关于做好高等学校家庭经济困难学生认定工作的指导意见》两份文件。^[2]但是, 与文件初衷背离的客观现象依然存在: 一方面, 很多来自贫困山区的学生, 由于认识上偏差而不愿申报, 导致贫困生数据覆盖率偏低; 另一方面, 在进行贫困审核时, 高职院校主要依赖生源地民政部门提供的证明。由于信息闭塞、区域数据量大, 很难进行实地考察等外部原因, 采集到的数据容易失真。因此, 承担具体工作的学生处往往很难做出较为全面客观的分析, 无法实现精准资助。

1.2 认定工作不全面, 资助偏离初衷

目前高职院校都建立了“奖、助、贷、减、免、补”的资助体系, 基本能满足所有贫困生不同层次的资助需求。但从这一资助体系我们也能看出, 当前的资助内容基本是满足贫困生的学杂费以及生活费的经济性补助, 较少体现心理辅导、人文关爱、价值观引导等深层次的资助内容。其实, 高职院校贫困生资助的初衷是在给予物质资助的同时也要深入育人层面。然而, 高职院校贫困生认定工作更多的是偏重经济因素, 很少考虑贫困生个体的学业、心理以及消费习惯等因素, 这些都是偏离资助初衷的。如果说, 专项资助资金的发放操作相对简单, 进一步完善就会减少真假贫困生混杂的场面, 杜绝“伪贫困生”现象, 那么改善贫困生认定工作中育人层面所涉及的心理辅导、人文关爱、学业跟踪等内容, 则需要教育工作者根据学生的性格、心理、学习

收稿日期: 2019-12-29

基金项目: 江苏省高等学校大学生创新创业训练计划项目(201912703026Y); 江苏省高等教育学会辅导员工作研究委员会专项重点课题(18FYHZD012); 2020 年南通市青少年工作研究课题(思想引领 20200202)

作者简介: 袁璟瑾(1982—), 女, 江苏南通人, 江苏航运职业技术学院学生工作处助理研究员, 硕士。

等情况给予长时间的关注、细致的把握,逐渐形成一个指标体系,建立一个全面合理的数据库,服务于贫困生筛选认定。

1.3 跟踪管理不及时,资助缺乏成效

高职院校贫困生认定工作一般一年一次,贫困生资格一旦认定下来,相关部门便建档立卡,审核工作也就此启动展开,这种流程安排有助于保证资助体系的连续性和稳定性,但同时也带来一定的弊端。譬如,个别贫困生为了享受国家资助,刻意隐瞒家中经济好转的状况,以过往的家庭情况申请贫困生资格。由于学校缺乏相应的追踪能力和手段,难以获得申请者的真实家庭情况,客观上导致资助资源不能精准资助实际更有需要的贫困生。同时,如有申请者由于家庭突发状况致贫,确实需要一定资助,然而由于申请流程复杂,一级级审核处理速度缓慢,临时再申请的贫困生很难获得及时资助。显然,高职院校迫切需要及时跟踪管理贫困生的各项数据,但多数高职院校处理相关数据依旧停留于人工操作,或者依靠国家和省级系统进行数据管理,校方自身极少具备数据信息化平台,这也使得贫困生资助缺乏动态管理,资助工作效率降低。

2 高职院校贫困生精准认定工作困境的成因分析

2.1 认定标准单一,标签效应负面

在贫困生认定的实际流程中,民政部门提供的证明几乎是唯一的依据。大多数情况下,高职院校要求学生自主提供相关证明材料,事实上放任的自助形式给部分申请者提供了做假机会。同时,在资助金发放过程中,高校为了保护家庭经济困难学生的“隐私”和“自尊”,一般不会公开贫困生的资料,无意中助长了“伪贫困生”的产生。更值得注意的是,高职院校实施的资助更多的是无偿的,只要达到国家资助基本标准就可以申请,当经济困难学生贴上“贫困生”的心理标签时,就会做出自我印象管理,使自己的行为与所贴的标签内容相一致。于是,有的贫困生出于所谓的面子原因,不愿接受院校安排的勤工俭学岗位,而有的贫困生产生“等、靠、要”的心理,认为他们享受这种无偿资助的福利理所当然。贫困生认定标准单一加上标签效应造成的种种乱象,既违背了国家资助贫困生的初衷,也违背了学校育人的本质。

2.2 数据整合乏力,动态管理缺失

随着智慧校园建设逐步展开,高职院校信息化平台得以建立。在此基础上,对学生在校期间的轨迹数据理应进行整合,对受资助学生的流水账、消费记录、校园一卡通等数据进行全面追踪和分析,对资助对象实施动态管理。但由于平台与平台的数据共享还未完全畅通,数据整合需要的指标体系尚未匹配,全面及时掌握学生基本信息的动态管理基础尚不完善,以至于资助过程中的大量数据还需依靠人工采集后输入,无法保证数据的时效性,也无法避免人工方式造成的漏报漏输、错报错输。

2.3 数据统计滥用,个体考评不足

高职院校的精准资助体系目前还缺乏整体性框架式设计,在收集使用贫困生数据时,无法应用多层次、多元化的有效数据,甚至出现严重偏颇。譬如,南京某高校利用学生一卡通在食堂消费的数据信息,实行对学生的饭卡消费补助,镇江某高校借助话费使用的大数据进行精准资助,应该说这些利用数据说话的做法值得肯定,但照搬照抄地套用在贫困生精准认定工作上就会衍生出系列问题,因为一卡通消费的多少不能完全反应学生的消费能力、话费的高低也无法完全确定一个学生是否贫困。可见,如果对大数据的认知只是停留在表面,滥用统计数据,使得数据说了假话,就会有悖于想让数据说话的初衷。

我们也要注意数据本质上具有不完整性,数据无法反馈申请者的个体性格,无法分辨申请者是否甚至谎报自身经济情况,是否有意占用社会公共资源。预防和根除“伪贫困生”现象需要其他的个体考评机制,不能完全依赖数据本身。

3 基于数据分析的贫困生精准认定工作框架构建思路

3.1 提炼指标权重,推进资助对象认定精准化

指标体系的构建有利于为大数据甄别有效数据,提供筛选分析功能,为教育管理者提供相应参考。^[3]高职院校可以结合国家政策和院校实际情况,制定一个基于大数据的客观评价指标体系。首先是提炼指标。资助管理部门要对录入系统的贫困生数据字段综合分析、提取,综合专家意见后,形成核定贫困生的指标,然后是明确指标权重。由于生源地域、家庭情况、院系指标、个性指标等原因,产生了多维度多元化的数据,根

据相关性、合理性来明确每一项数据在指标权重中的占比,最后为构建立体化指标体系,用来显示贫困生受资助以来的成长轨迹。通过贫困生个体的轨迹大数据确定贫困生更需要的资助方式,选择是经济性资助还是发展性育人资助,从而给予学生更精准的资助类型。

3.2 分类区别管理,推进资助行为差异化

贫困生致贫原因很复杂,需要分类建卡立档以准确反映贫困生现状。学校可以利用数据分析中模糊C均值聚类法将学生数据并行分类,按照贫困等级、致贫性质、贫困生心理状态进行分组,并利用每个组群高度的相似性,进一步实现贫困生资助工作精细化。通过规范贫困生数据档案和完成贫困生档案智能化的基础工作,学校就可利用学校大数据池的数据,将不同技能、专业、特长的学生安排至学校不同勤工俭学岗位,减轻学生的经济负担,也让学生所学专长得以发挥,实现真正意义上的精准资助脱贫。

3.3 善用信息化技术,推进资助管理动态化

通过信息化技术构建大数据精准资助系统是推进资助管理动态化的必要工作。高职院校可以利用智慧校园的建设,有序建立学生信息“大数据池”,全面存储学生基本信息及相关轨迹数据。基本信息不仅涵盖学生学籍信息、户籍信息、家庭成员信息等,也要包括学生缴费记录、受学院资助状况、贫困生认定证明材料、校园一卡通消费记录、银行卡异动及其购买记录等。在收集足够数量的大数据的基础上,进行数据清理、提取、分类和梳理,让数据真实全面地反映学生的学习、生活情况,实时监测数据进行预警,推进资助数据动态化管理,让急需资助的学生得到及时的帮助。同时,结合民主测评、调查访谈、一对一帮扶等途径,多渠道掌握受资助学生的实际情况和发展动态,在实现精准资助的基础上,给予更多的人文关怀。

3.4 构建科学考评标准,推进资助考核合理化

资助工作绩效考核意义重大,从某种程度上说,是成果验收,也是正向驱动,主要基于资助基本工作、实施成效、育人功能等多项指标综合考量。一个科学完善的考核标准应该涵盖但不限于以下内容:受资助学生的学业成长、生活质量提升程度、资助制度的贯彻落实等。在具体应用中,可以在利用大数据对贫困生精准识别、个体差异化资助、动态管理的基础上,实现过程化考评,利用过程数据说话;也可以拓宽评价主体,监管主体拓展为学生、教师、家长、社会等,建立多元监管;还可以运用人力资源中常用的AHP方法,建立评价模型,通过层次分析法明确考核指标体系,为过程中每一层次的考核提供参考依据,最终实现考核合理化及精准化。

4 基于大数据的精准资助实施对策

4.1 构建符合大数据背景的高校资助对象精准认定指标体系

根据前期的指标提炼、权重分析,在大数据背景下构建基于学生消费、学习等的三级指标体系,如表1所示。由于学生消费水平是初步判定贫困等级的基础,所以指标体系重点分析学生消费行为。在个人消费情况的一级指标里设置3个二级指标、20个三级指标,反映学生在校内整体的一个消费水平,为后期贫困生精准认定提供参考。资助的最终目的是要实现育人功效,在学习情况一级指标里设置了2个二级指标、13个三级指标,反映学生整体学业情况以及日常生活。受资助状态指标可以设置参数值,受资助的用1表示,未受资助的用0表示。生源地指标可以设置三种参数,在数据权重比中体现地区差异。

4.2 精准资助相关数据预处理和数据清洗

通过“数据池”共享集成的所需数据库进行数据预处理,首先是对提炼处理指标的每个数据表进行整理分析,其次是将数据表集成,最后是检验冲突数值。例如,在图书馆门禁系统信息和宿舍门禁系统信息中包含同一个学生的N条刷卡记录,这往往是重复刷卡造成的数据冗余,我们需要去除重复数据,保留最后一条数据记录,以便准确反馈学生的正确信息;在整理在校消费记录时,可以用数据库SELECT语句统计,如select student-id, count(*) from card where 消费='POS' group by student-id;在图书借阅统计中将不同类的书用首字母分类,同时用COUNT函数在数据库中统计分析等等。在宿舍门禁系统信息中,我们可以通过数据库中中学生ID、进门时间、出门时间三个变量统计出学生进出宿舍的总次数以及早上7点50前出宿舍总次数、晚上9点30之后回宿舍总次数。

表 1 高职院校资助对象精准认定指标体系

一级指标	二级指标	三级指标
个人消费情况	一卡通消费	学生一卡通 ID
		校园消费总额
		校园消费总次数
		校园消费均值
		一卡通剩余金额最大值
		一卡通剩余金额最小值
		一卡通剩余金额平均值
		一卡通剩余金额中位数
	日常生活消费	学院超市消费总金额
		学院超市消费总次数
		学院超市消费平均值
		打开水消费总金额
		打开水次数
		学院浴室消费总金额
	餐厅消费特征	学院淋浴次数
		学院餐厅用餐总金额
		学院餐厅总次数
		平均用一餐金额
		单次消费最大金额
		单次消费最小金额
学习情况	学习成绩	专业排名
	日常表现	进图书馆总次数
		图书馆借书总量
		借书种类
		借专业书籍总量
		借人文社科类总量
		借自然科学类总量
		借军事类总量
		借综合类图书总量
		进宿舍总次数
		出宿舍次数
		早上 7 点 50 前出宿舍总次数
		晚上 9 点 30 之后回宿舍总次数
受资助状态	是否取得助学金	
生源地	东部	
	中部	
	西部	

对数据预处理结束后进行数据清理环节,为不影响实验结果,采用基于统计的检测方式或基于聚类离群点的检测方法进行异常数据的删除。首先,删除标准差相差较大的、呈现两极分化较严重的数据;其次,删除离散变量缺失比例超过 70%、无法插补填充的数据;最后,对于正态分布的连续变量数据,均采用平均值法填补数据。

4.3 采用 SMOTE 方法实现资助对象精准认定

在所有学生数据中,受资助数据样本较少,直接运用分类模型进行受助对象认定往往准确率不高。采用 SMOTE 采样平衡数据,可以提高分类效率,准确识别受资助者的特征。本次实验采样 10 236 条样本数据,其中受资助样本 1 535 条,通过 SMOTE 重采样方法,选取 5 个近邻进行计算。计算数据集如表 2 所示。实验结果显示,SMOTE 处理后的数据不平衡率显著改善,较大提高了受资助者认定的准确率,有助于后期建模预测的精准度将大幅度提升,也能准确地识别受资助学生的特征,对于贫困生精准认定可以起到很好的效果。

表 2 计算数据集示例

数据类别	样本数据(10 236)		SMOTE 处理后的数据(14 913)	
	样本总数	不平衡率/%	样本总数	不平衡率/%
未受资助	8 701	85	8 948	60
受资助	1 535	15	5 965	40

5 结束语

从当前贫困生资助工作面临的困境及其成因分析可以看出,下一阶段高职院校贫困生资助工作的关键在于提高精准度,本研究基于数据分析构建了贫困生精准认定框架。实现真正意义上的贫困生精准资助需要利用好大数据,更需要进一步完善精准资助的联动机制,不仅在经济方面为学生提供方便,更要能在发展性育人方面实施精准资助,这是资助体系发展完备的需要,也是国家资助贫困生政策实现扶贫育人的初衷。

参考文献:

- [1]习近平.习近平谈扶贫[EB/OL].(2016-09-01)[2020-02-20].<http://theory.people.com.cn/n1/2016/0901/c49150-28682345.html>.
- [2]刘晓杰.“精准扶贫”思想下的大学生“精准资助”[J].教育教学论坛,2017(3):3-5.
- [3]张名艳,张阳.论精准扶贫视域下高校贫困大学生的就业精准帮扶[J].教育与职业,2019(8):63-65.

(责任编辑 顾力豪)

Exploratory Analysis of Accurate Verification of Students in Poverty in Higher Vocational Colleges Based on Big Data

YUAN Jing-jin¹, KONG Shu-pe², CHEN Yan²

(1. Department of Student Work, Jiangsu Shipping College, Nantong 226010, China;

2. School of Transport Management and Economics, Jiangsu Shipping College, Nantong 226010, China)

Abstract: The reliance of big data technology on realizing accurate financial aid for students in poverty in higher vocational colleges is the trend for the reform of future students' financial aid management. Presently the accurate financial aid in higher vocational colleges exists the problems in data collection, the verification and tracking management. They are mainly caused by unitary evaluation standard and negative label effect, lack of data consolidation and dynamic management, abused data statistics and insufficient individual evaluation. It is proposed to establish a working framework of accurate verification of students in poverty based on big data, promote the precision of verification for aided individuals, the differentiation of aid behaviors and the rationalization of the aid assessment. Finally, it is suggested to adopt the method of SMOTE to accurately verify individuals to be aided, and substantially improve the effectiveness of the accurate aid.

Key words: big data; student in poverty; accurate verification