

大数据背景下高校贫困生精准资助工作模式研究

张晓双

青岛理工大学 山东 青岛 266033

【摘要】信息数字化时代的到来,为高校的贫困资助方式提供新的理论与技术的支持,同时也让各个高校找到新的机遇研究精准资助工作。我国绝大多数高校致力于校园信息化网络建设中,本次研究从研究学生在学校的日常行为入手,分析学生图书借阅数据、图书馆门禁数据、宿舍门禁数据、成绩排名数据、校园一卡通消费等。从各个方面综合且全面的考虑学生。避免了“假贫困”问题,能够对贫困学生助学金精准提高。在大数据的时代背景下,本文基于机器学习方法,使用校园一卡通数据对助学金的发放进行精准的预测。

【关键词】大数据;高校;贫困生;资助

随着科技水平的提升、数字化、信息化的时代到来,目前我国各个高校都大力开发校园信息化建设,而校园一卡通是其重要的组成部分,校园一卡通记录着学生的个人信息,因此具有个人身份认证、金融服务功能、综合消费的功能等等。校园一卡通集成学校所有资源,通过校园一卡通数据的采集能够充分掌握学生的动态、实时的消费情况。一卡通的用途不仅要用于消费结算的功能而是通过智能的数字化管理对学生生活的一个真实写照。

为帮助在校的普通本科高校和高等职业学校的家庭经济困难学生顺利完成学业,国家施行助学金政策。近几年国家大力发展教育,鼓励学校扩招,如何对助学金的精准资助,成为我们关注的问题,助学金精准资助方法研究也越来越受到大家的关注。本次研究从研究学生在学校的日常行为入手,分析学生图书借阅数据、图书馆门禁数据、宿舍门禁数据、成绩排名数据、校园一卡通消费等。从各个方面综合且全面的考量学生。避免了“假贫困”问题,能够对贫困学生助学金精准提高。

1 现状分析

我国现在对于高校大学生的资助准确性存在的问题如下:

对于学生认定方式单一。目前我国高校多采用学生家庭情况调查——资助申请——贫困生审核——投票选出结果,这种情况表面公平实则很多细节需要仔细调查。

资助后对学生的引导不到位。在高校评奖学金的过程中最看重的是成绩,通过考试成绩排名来确定奖学金,这就造成资助后引导存在问题。不能达到良好的扶贫效果。

资助体制存在问题。在功效的资助体制中,有时会出现资助不足或者重复资助相同的学生,资助的同时部分学生会助学金产生依赖性。

资助制度存在问题。第一,学生填报信息出现虚假情况,第二,资助后对贫困生的指导和思想教育存在不足。

2 相关技术

数据描述。实验数据根据学生的日常分为三类:图书馆借阅书籍数据、校园卡生活消费数据、学生历年平均成绩数据。

特征构造。从介绍的五大类数据进行特征提取。所有特征表示成向量: \mathbf{x} , N 为特征维数。本文特征维数为 555。由于数据中用户天数存在较大偏差,所以用户消费总额或者计数等指标应除以用户的有效天数,才能保证特征更加具有“公平性”。

特征处理。由于特征中含有部分缺失值,因此对特征进行缺失值填充处理。缺失值填充的方式有很多种:均值、众数、中位数、0、-1 进行缺失值填充。

特征选择。特征选择有三种方法:过滤型、嵌入型、包裹型。结合本文对三种方法的比较,本次选择包裹型方法对特征进行选择。

模型选择。机器学习模型分为树模型和非树模型。本文使用目前最常见的三种树模型:Random Forest、GBDT、XGB。

3 实验结果

本次研究实验数据采用某高校学生的校园一卡通分消费数据。研究对象共有 12231 名学生,其中没有获得助学金与奖学金的人数共有 9654 人,占本次研究对象人数的 84%;获得三等助学金的有 782 人,占本次研究对象人数的 7%;获得二等助学金的有 554 人,占本次研究对象人数的 6%;获得一等助学金的有 395 人,占本次研究对象人数的 3%。本次实验的样本数据存在分布不平衡性,所以在对数据进行处理时要进行保证样本权重相等。采用两种方法:数据采样和设置样本权重。

分析得出学生成绩的排名、校园卡消费数据,尤其是食堂消费数据和地点成为对是否获得助学金的最大影响因素之一。由此分析三个模型的效率: $GBDT < RF < XGB$ 。因此将最后的结果按照分类器的性能进行加权平均得到最后的结果为 0.02943。模型融合后的结果比 GBDT 的结果高了 0.00770、比 XGB 的结果提高了 0.00222、比 RF 的结果提高了 0.00106。

随着高校进入大数据时代,传统的助学金预测方法已经存在许多问题,因此如何从海量的学生一卡通数据中提取有价值的特征,进而预测用户获得助学金情况成为现代高校的重要研究内容。本文以高校学生数据为基础,探讨了用户特征提取方法,利用监督机器学习方法,多模型加权融合,预测用户获得助学金概率。最后,笔者基于上述方法开展了助学金预测的检验工作。在未来的工作中,笔者将继续完善预测模型,并将其优化。通过扫描学生的一卡通,可以显示出该学生获得助学金或者奖学金的概率以及需要改进的地方。这个设计一方面可以激励身边的同学学习也可以让学生自己控制自己的消费水平。

参考文献:

- [1]郭瑞玲,杨景.基于“互联网+”思维下精准帮扶在班级建设中的应用[J].科技风,2019(30):107.
- [2]辛翔,黄厚华,贾迅.基于大数据视角下高校精准资助面临的问题及其解决路径[J].智库时代,2019(43):115-116.
- [3]高黎.高校辅导员视角下的大学生精准资助工作创新研究[J].创新创业理论与实践,2019,2(18):157-158.
- [4]顾小丽.“人工智能+”贫困生精准识别工作探析[J].科技创新与生产力,2019(09):44-46+49.