

下面开始国内外文献分析，分析利弊以及精度。下面需要你对这些篇参考文献进行总结，我会给你你所列出的第一条参考文献及其摘要我总结的例子，请你以此为模板将剩余的参考文献进行总结：

[1]肖励. 银行客户购买产品意向预测模型研究[D]. 中南财经政法大学, 2020.

随着“互联网+”的快速发展和大数据技术应用的不断深入,各行业积累了大量的用户行为信息。基于更好满足用户需求,以用户为中心的行为分析成为各行业的焦点,提供良好的用户体验服务是企业发展的关键。传统商业银行业务体量大、用户数众多,银行期望通过对用户历史行为做分析,预测用户未来会增加购买的银行产品,以便根据用户的需求调整自身的业务,适应互联网金融发展的潮流。本文研究分析银行用户的历史行为数据,提出一种银行客户购买产品意向预测模型。首先,对数据进行特征工程处理,通过数据清洗、分类和筛选,构建本文的输入特征向量。其次,采用支持向量机(SVM)、最小二乘支持向量机(LS-SVM)以及对比模型 XGBoost 和 LightGBM 等模型在数据集上进行训练,并使用带交叉验证的网格遍历法对模型参数做优化,分别针对惩罚因子和核函数开展参数调整和优化工作,寻找最优的参数组合,得到模型的较好的效果。最后,将支持向量机模型同 XGBoost 模型、LightGBM 模型等其他模型的性能效果作对比分析。实证研究采用西班牙桑坦德银行的公开数据集。采用数据统计分析技术,进行数据特点分类统计,分析原始数据的隐含信息。基于数据统计分析结构,再通过特征工程对数据进行特征提取,从初始特征、简单特征和复杂特征等三个层面分析构建本文的数据特征。进一步,运用模型进行银行客户购买倾向预测,根据用户前 17 个月购买产品的历史行为记录,预测当月用户新增购买产品的情况。银行客户可能会新增加购买多个银行产品,本文将原始问题转化为一个多分类问题,依据预测用户购买各类产品的概率进行排序,选择概率最大的多个产品有序地推荐给用户。最后,采用通用的评估指标对模型效果进行性能评估。研究发现,本文得出最优的组合参数,采用线性核函数的 LS-SVM 取得最好的效果,分类准确率达到 87.41%。本文提出银行客户购买产品意向预测模型可适用在国内各大银行产品业务上,从而提高银行工作效率和业务水平。本文的创新点和贡献主要有:一是

提出一种融合的特征提取框架和方法,采用特征工程对原始特征进行清洗、分类和重构,形成了新的特征,该特征提取方法可以较好和有效的表达银行客户的历史购买行为的特点。二是构建一种融合支持向量机的预测模型进行购买产品的多分类问题预测。在构建最优参数模型的过程中,采用网格遍历法对支持向量机进行参数的优化,训练得到最优的预测模型,再用于客户未来购买产品意向预测。本文工作对银行客户未来购买产品意向预测研究具有重要参考价值,为银行向客户推荐理财产品提供决策依据。

”肖励等人利用统计分析技术以及支持向量机、最小二乘支持向量机、XGBoost 以及 LightGBM,对银行客户购买产品意向数据进行了很好的数学建模,并详细介绍了其特征工程、模型训练以及模型对比分析,采用线性核函数的 LS-SVM 取得最好的效果,分类准确率达到 87.41%,但是对模型的能力挖掘不足以及对数据的正负样本不均衡挖掘不足。“

上面是例子请你分别对你所找到的剩余文献的”摘要部分“进行分析与汇总。直到输出完整才停止。每篇论文的精度是十分重要的,就像肖励论文中的分类准确率达到 87.31,每篇必须要有一个数据进行支持.

[1]谢青芳. 基于机器学习的用户购买行为预测[D]. 兰州大学, 2023. DOI:10.27204/d.cnki.glzhu.2023.003614.

随着互联网的飞速发展,电商消费者规模和商品规模不断扩大,电子商务在为用户购物提供便利的同时,也带来了信息过载问题.基于购买预测,可以设计个性化的商品推荐和营销方案,有助于缓解信息过载、优化用户购买体验.电商平台海量的用户交互行为数据背后隐含了用户对商品的偏好,因此本文基于这类隐式反馈数据构建购买预测模型,对用户的购买行为进行预测.本文的主要工作和创新点体现在以下几个方面.在特征构造方面,隐式反馈数据的特征维度有限,并且以往的研究大多基于单一的时间窗口.本文基于电商场景中用户与商品的真实交互记录,利用时间滑动窗口方法构造特征,并设置了长期、中期、短期三种不同长度的时间窗口,以挖掘用户不同时期的兴趣偏好.在每个时间窗口内从用户、商品、品类、用户-商品、用户-品类五个维度进行特征构造.最终共构建了 84 个特征,有效地扩充了隐式反馈数据的

特征维度,为特征构造提供了思路.在模型构建方面,本文使用了 XGBoost、LightGBM 和 Cat Boost 这三个前沿的机器学习算法进行用户购买预测,结果显示,LightGBM 模型的 F1 值最高,在不平衡数据上有更好的表现.为进一步提升预测效果,本文构建了一种基于参数扰动的同质融合模型,该模型使用 Bagging 方法融合若干个 LightGBM,并对基学习器进行参数扰动,进一步增大各基学习器间的差异.实验结果显示,基于参数扰动的同质融合模型 F1 值达到了 0.6107,比单一模型的预测效果更好.

[2]陈龙. 基于机器学习方法的用户复购行为预测[D]. 南开大学, 2021. DOI:10. 27254/d. cnki. gnkau. 2021. 000097.

随着互联网的快速发展与普及,网购者规模迅速扩大,电子商务迅猛发展,这也导致各大电商竞争愈演愈烈。近几年电商会进行大促活动,如双十一,六一八等发放优惠券以吸引新的消费者,但是此类活动效果有限,没有带来长期回报,因为很多消费者只是一次性消费者。为了取得更高的回报,商家们需要研究出哪类消费者才是值得为之付出的重复购买者。因此,研究消费者的复购行为有很强的应用价值。本文根据天猫 2017 年“双十一”用户行为记录数据对网站的复购行为进行了探究。数据包括 11.11 当天一些商户的新客户在活动日及之前 6 个月内的信息,首先本文从业务角度通过分析影响重复购买的因素设计特征工程提取特征,并使用了三种模型进行建模以及对应的模型融合,其算法在实验中性能表现不错。本文的工作主要包含以下几个方面:(1)数据处理。通过数据简介、数据探索、数据预处理,更好地了解数据集的字段含义及其与标签是否复购的关系。(2)设计特征工程。通过分析影响重复购买行为的因素,提取出用户特征、商家特征、用户与商家的交互特征这三个角度的特征;在升维之后进行了特征选择。(3)研究单一模型与融合模型在特征工程上的运用。分别使用 XGB、LGB、catboost 进行建模,最后几个模型结合在一起,组合模型能互补各个模型的优势获得更好的预测结果。

[3]梁宇婷. 基于过采样技术与机器学习的用户复购行为预测模型[D]. 重庆大学, 2021. DOI:10.27670/d.cnki.gcqdu.2021.001532.

随着电子商务的发展与大数据时代的到来,传统的营销策略已经不能满足电商的快速发展,电商越来越倾向于从海量消费者数据中挖掘用户偏好、发掘潜在客户,并提出更为客观的数字营销战略。本文基于用户画像、行为和购买等信息构建相关的用户复购预测模型,通过实证分析与比较发现基于 Borderline-SMOTE 过采样技术的 Voting 与 Stacking 融合模型的预测效果与稳定性最优。为降低商家的库存与营销等成本,对用户进行精准营销,从而有效提高投资回报率(ROI)有一定的借鉴意义。本文的主要研究工作如下:(1)首先,本文根据天猫用户画像和消费数据进行可视化分析,大致了解了整体复购情况和用户性别与年龄对复购的影响情况。(2)根据原始变量主观构建了用户、商户和用户商户交互三个层面且与目标变量相关的新特征。并运用 word2vec 中的 CBOW 模型对商户 ID 进行编码。为了剔除冗余与无效特征,运用标准差筛选法和极端随机树筛选法对所有特征变量进行筛选,最终筛选出离散程度低,与用户复购行为密切相关且对模型预测贡献程度最高的 36 个特征。(3)由于在所有用户中复购者仅占 6.12%,目标变量呈现不平衡分布。本文运用 Borderline-SMOTE 过采样方法对数据进行随机采样,很好地解决了目标变量不平衡分布对模型预测的影响。(4)根据过采样后的数据构建基于 K 近邻、随机森林与 Cat Boost 的单一模型,并运用交叉验证和随机与网格搜索方法不断迭代模型对参数进行优化,对比预测效果发现 Cat Boost 预测模型效果最好,在测试集上的预测准确度达到 95.44%。以单一模型为基模型进一步构建 Voting 与 Stacking 融合模型,经过实证分析比较,加权投票与 Stacking 融合模型在训练集与测试集上的预测准确率和 F1 值均达到 96%以上,相对单一模型具有更强的稳定性与泛化性能。通过预测效果对比分析,基于 K 近邻、随机森林与 Cat Boost 的 Stacking 融合模型对于用户复

购行为预测效果最好,很好地提升了单一模型的泛化能力与稳定性。为电商识别忠诚客户进行有效推广有很好的模型借鉴意义。

[4]叶艳艳. 基于机器学习的电信用户流失预警模型预测与分析[D]. 山东大学, 2021. DOI:10. 27272/d. cnki. gshdu. 2021. 003537.

在日新月异、高速更迭的信息化时代中,电信用户在商品、内容、服务方面都有着新的消费诉求.当前电信行业所有的收入导向的运营举措、单一的套餐营销策略、缺乏个性化的服务等已经难以充分满足客户需求.在市场接近饱和,行业间竞争主要在于用户的留存,收入与客户满意程度“两难全”的大背景下,电信行业需要快速灵活地应对市场变化,降低本网流失概率,抢占异网流失用户市场.因此,结合本人实际在某运营商的实习经验以及前人的大量研究,本文提出了分群构建电信用户流失预警模型.在前人预测流失模型的基础上,更加重视依据用户群特征,分析用户流失倾向及原因.首先对 100000 个某运营商用户,共 99 列用户特征数据进行数据清洗,经过 Pearson 相关性检验后保留 66 个特征变量.剩余变量被分成使用、费用和基础信息这三个维度,对每个维度创新地采用基于机器学习的嵌入法进行特征筛选,选出三个维度,共 18 个指标.对筛选后的每个维度的用户聚类分群,采用 KMeans 聚类算法分群的方式,将每个维度各分成 2 类,精准显示消费水平,语音使用水平,活跃程度等信息.使用 PCA 降维完成用户分群,最终将用户分为 8 类.同时通过信息规整,对全部用户群体进行典型画像描述,便于管理人员和运营人员更加知晓用户行为习惯和使用偏好信息.其次开始分群预测流失概率,使用前人预测用户流失较好的机器学习算法,包括随机森林,Adaboost,GBDT,Extra Tree,XGBoost 以及创新地使用了以软投票分类器为融合机制的投票表决法.通过对所有用户的模型预测结果的比较,选择了以软投票为形式的投票表决法,其 F1 得分为 71.8%.针对不同特点的用户群分类构建模型,预测用户流失倾向标签,比较分类前后的模型 F1 得分分别为 71.8%和 75.4%,证明了分群有效.为了提高模型效果,融合前人的经验,采用 SMOTE 算法平

衡正负样本.经过样本平衡后的 F1 得分最高提升了 6.04%,从而进一步更好地预测用户流失倾向.最后利用不同类用户的特征以及影响用户流失倾向的重要因子系数排序,对用户流失原因进行分析,给出用户留存建议以及对应决策.在对客户分群知悉的基础上,把握用户个性化消费需求,以做到更细致、更精准的营销推荐.

[5]侯荣真. 基于不平衡数据集的机器学习算法对用户流失的预测[D]. 南开大学, 2021. DOI:10. 27254/d. cnki. gnkau. 2021. 000070.

随着互联网的高速发展,音乐流媒体不断壮大,各大音乐平台通过多种方式抢占市场份额。一般来说,在市场高度饱和的情况下,留住老用户的难度低于吸引新用户,因此客户关系管理中的用户流失预测成为了企业管理的关键一步。同时,海量数据的收集与存储为机器学习技术提供了便利,因此利用机器学习算法进行客户流失预测成为了主流方式。本文即是基于机器学习算法,对音乐软件 KKBOX 进行用户流失预测。但是数据的庞杂性和类别不平衡性是用户流失预测的难点,因此在大量数据中进行合适的指标选取与模型选择成为了本文的研究重点。论文的主要工作如下:(1)首先在对数据有了基本了解后,通过时间窗口加原始数据变量的方式进行新特征的构造,之后再基于嵌入法进行特征筛选,最终选出 12 个对用户流失预测较为重要的变量,如 `is _ auto _ renew _ 60`,其代表最近 60 天内自动续订会员的次数,是由 60 天的时间窗口加上原始变量 `is _ auto _ renew` 结合而来的。(2)利用 Xgboost、Light GBM、随机森林三种集成学习算法结合随机欠采样、Easy Ensemble、SMOTE 过采样进行流失用户的预测并对比其预测效果。发现基于 Easy Ensemble 的 Light GBM 算法效果最优,其 AUC 值和 Recall 值为 9 种组合方式中最高的。(3)在上述 9 种组合方式研究的基础上进行模型探索,提出了将 Borderline-SMOTE、ENN、Easy Ensemble 相结合的采样方式,并用 Light GBM 算法进行预测。将其与基于 Easy Ensemble 的 Light GBM 算法进行对比,发现此种方式可以使 AUC 值保持在一定水平的基础上,提升模型的 Recall 值。因此

若想要在不过多提升代价的情况下,尽可能多的识别流失用户,此种方式具有一定的可行性。

[6] 胡雅竣. 基于机器学习的银行用户信用违约预测方法研究[D]. 华中科技大学, 2021. DOI:10.27157/d.cnki.ghzku.2021.002557.

信用支付业务已经越来越重要的成为我国银行业和金融业的重要业务,随着社会现代化的发展,具有越来越多样的载体形式和产品形态,如何高效精准的进行银行用户信用违约预测成为了极其重要的问题。针对信用支付业务场景,本文基于机器学习方法设计了银行用户信用违约预测方法。研究了多维度用户违约特征构建方法来解决银行用户信用违约预测特征构建不充分问题;建立了基于分位点的归一化方法来解决银行信用违约数据分布不均衡问题;建立了基于逻辑回归的前向特征筛选算法来解决银行信用违约预测模型构建中特征空间膨胀的问题;建立了线性加权模型融合方法进行模型融合来解决单模型预测偏差与精度问题。为验证本文所提方法的有效性,参加了 2020 年某银行 Fintech 挑战赛,并取得了决赛第三名的成绩。通过对比赛结果进行分析,进一步优化了算法。在某银行用户信用违约数据集上的测试结果表明:多维度用户违约特征构建方法将三种不同类型机器学习模型的 auc 值平均提高了 7.21%;基于分位点的归一化方法将逻辑回归模型 auc 值提高了 0.72%而不降低 xgboost 和随机森林模型的表现;基于逻辑回归的前向特征筛选算法将三种不同类型机器学习模型的 auc 值平均提高了 0.15%同时将模型训练速度加速了 3-5 倍;线性加权模型融合方法得到的融合模型相较于单模型最优 auc 值提高了 0.10%。实验结果进一步验证了本文针对银行用户信用违约问题所研究建立的建模方法的有效性和合理性。通过对本问题的研究可以发现,一个高效准确的机器学习模型,往往需要在目标业务场景进行合理详细的背景调研和数据分析的基础上,结合所使用机器学习算法的特点,对目标业务难点进行有效的针对性优化,综合

考虑不同机器学习建模方法的优势和劣势,在偏差与方差、模型复杂度和可解释性、优度和速度之间寻找平衡点,才能满足特定业务场景下的要求。

[7]张晨楠. 基于机器学习的用户贷款风险预测模型研究[D]. 海南大学, 2020. DOI:10. 27073/d. cnki. ghadu. 2020. 000268.

随着借贷平台的兴起以及互联网金融的发展,金融行业的用户贷款风险预测也变得更加的重要,传统的金融行业面对数据量的剧增已经审核成本的提升,已经难以进行高效的数据处理。随着计算机技术的发展,在大数据时代机器学习技术的出现,给我们提供了更多的可能性和便利性,面对大量的借贷用户,网络借贷平台也出现了一系列的风险管理问题,导致平台方和用户方的合法权益受到损害,因此相关部门也出台了政策来进行约束管理来促进贷款平台的发展。对于借贷平台,自身也该采用技术手段来进行风险规避,故而采用机器学习技术建立预测模型提取出有效信息,进行风险预测来有效的控制风险并最大限度的减少损失。本文针对用户贷款风险预测问题,探讨机器学习方法在互联网金融领域的应用情况。本文研究是在基于以往研究者的基础上采用某互联网贷款平台提供的经过脱敏处理后的用户贷款相关的数据集,本文的主要研究内容如下:(1)首先对用户数据进行了数据预处理的操作,对于与用户的相关个人基本信息和用户相应的贷款相关的信息数据进行了探索性分析,对数据集进行了数据预处理和清洗,包括对缺失值以及重复值进行处理,并对时间戳进行了补齐;(2)在特征工程处理方面,重点对数据集的特征进行处理,采用特征交叉组合方法进行特征衍生、对部分特征进行 one-hot 编码、部分变量的归一化处理等,特征选择采用 Random Forest 算法按照特征重要性排序,选取重要性排名前 15 的特征,并完成最终输入模型的变量汇总;(3)构建模型以及模型的优化,将训练集分为测试集和验证集,把经过特征选择出来的新数据集作为 XGBoost 模型的输入,经过参数调优和交叉验证得到最终模型的最优参数,并在测试集上进行模型的应用。对融合后模型的性能进行评估,并将改进的预测模型与逻辑回归模型和



GBDT 模型的预测效果进行对比,得到新模型的预测效果明显是优于其他两个预测模型。通过实验本文提出了一种基于随机森林的 XGBoost 模型用于用户贷款的风险预测,该模型有较好的预测准确性,最终根据该模型得到的结果结合带大数据时代的背景,对网络借贷平台识别高风险的借贷用户提出一些具有参考性的建议。

[8]张韶. 基于机器学习的用户购买行为预测研究[D]. 长安大学, 2020. DOI:10. 26976/d. cnki. gchau. 2020. 000772.

随着互联网的飞速发展,电商消费者规模和商品规模不断扩大,电子商务在为用户购物提供便利的同时,也带来了信息过载问题.基于购买预测,可以设计个性化的商品推荐和营销方案,有助于缓解信息过载、优化用户购买体验.电商平台海量的用户交互行为数据背后隐含了用户对商品的偏好,因此本文基于这类隐式反馈数据构建购买预测模型,对用户的购买行为进行预测.本文的主要工作和创新点体现在以下几个方面.在特征构造方面,隐式反馈数据的特征维度有限,并且以往的研究大多基于单一的时间窗口.本文基于电商场景中用户与商品的真实交互记录,利用时间滑动窗口方法构造特征,并设置了长期、中期、短期三种不同长度的时间窗口,以挖掘用户不同时期的兴趣偏好.在每个时间窗口内从用户、商品、品类、用户-商品、用户-品类五个维度进行特征构造.最终共构建了 84 个特征,有效地扩充了隐式反馈数据的特征维度,为特征构造提供了思路.在模型构建方面,本文使用了 XGBoost、LightGBM 和 Cat Boost 这三个前沿的机器学习算法进行用户购买预测,结果显示,LightGBM 模型的 F1 值最高,在不平衡数据上有更好的表现.为进一步提升预测效果,本文构建了一种基于参数扰动的同质融合模型,该模型使用 Bagging 方法融合若干个 LightGBM,并对基学习器进行参数扰动,进一步增大各基学习器间的差异.实验结果显示,基于参数扰动的同质融合模型 F1 值达到了 0.6107,比单一模型的预测效果更好.

[9]Wenlong L .Default Prediction of Internet Finance Users Based on Imbalance-XGBoost[J]. Tehnički vjesnik, 2023, 30 (3) :779-786.

Fast and accurate identification of financial fraud is a challenge in Internet finance. Based on the characteristics of imbalanced distribution of Internet financial data, this paper integrates machine learning methods and Internet financial data to propose a prediction model for loan defaults, and proves its effectiveness and generalizability through empirical research. In this paper, we introduce a processing method (link processing method) for imbalance data based on the traditional early warning model. In this paper, we conduct experiments using the financial dataset of Lending Club platform and prove that our model is superior to XGBoost, NGBoost, Ada Boost, and GBDT in the prediction of default risk.

[10]Jing G ,Wenjun S ,Xin S .Research on Default Prediction for Credit Card Users Based on XGBoost-LSTM Model[J].Discrete Dynamics in Nature and Society,2021,2021

The credit card business has become an indispensable financial service for commercial banks. With the development of credit card business, commercial banks have achieved outstanding results in maintaining existing customers, tapping potential customers, and market share. During credit card operations, massive amounts of data in multiple dimensions—including basic customer information; billing, installment, and repayment information; transaction flows; and overdue records—are generated. Compared with preloan and postloan links, user default prediction of the on-loan link has a huge scale of data, which makes it difficult to identify signs of risk. With the recent growing maturity and practicality of technologies such as big data analysis and artificial intelligence, it has become possible to further mine and analyze massive amounts of transaction data. This

study mined and analyzed the transaction flow data that best reflected customer behavior. XGBoost, which is widely used in financial classification models, and Long-Short Term Memory (LSTM), which is widely used in time-series information, were selected for comparative research. The accuracy of the XGBoost model depends on the degree of expertise in feature extraction, while the LSTM algorithm can achieve higher accuracy without feature extraction. The resulting XGBoost-LSTM model showed good classification performance in default prediction. The results of this study can provide a reference for the application of deep learning algorithms in the field of finance.

[11] 常文晗. 基于 Adacost 算法的居民保险购买行为影响因素分析[D]. 天津财经大学, 2020. DOI:10.27354/d.cnki.gtcjy.2020.000752.

近些年来,随着人工智能技术的兴起,机器学习算法已经被广泛地应用到了各个领域并且取得了显著的研究成果,但是机器学习算法在我国保险行业尤其是在小型保险公司中的应用还不够成熟,同时,我国的小型保险公司也急需改变以往的营销模式,探索数字化营销模式。机器学习算法能很好地应用到这种营销模式中并且能够极大地降低保险公司的营销成本,提高其综合竞争力。在这样的背景下,本文主要研究了分类学习算法在保险行业中的客户分类以及影响因素分析等方面的应用。本文收集了可能影响居民保险购买行为的家庭因素与个人因素并以此进行了分析,得出模型特征的重要性与分类规则,为保险公司在产品营销以及业务拓展方面提供了一定的参考意见。本文首先对我国居民的保险购买行为进行了一定的阐述,介绍了我国保险业从起到到现在各个阶段内保险消费者的行为特征,讨论了可能影响我国居民保险购买行为的个人因素与家庭因素。本文使用 2016 年的 CFPS 数据,对数据进行建模并且比较了各类模型分类的准确率。计算结果表明:加入误分类代价后,Adacost 算法能够很

好地解决样本不平衡问题,对居民购买商业保险的行为预测较为准确,能有效识别有价值客户。根据模型分析的结果,得出模型特征重要性以及相应的分类规则,为保险公司在产品营销和业务拓展方面提出了相应的参考意见。

[12]祝歆,刘潇蔓,陈树广等. 基于机器学习融合算法的网络购买行为预测研究[J]. 统计与信息论坛, 2017, 32(12):94-100.

选择适合的机器学习算法是在社会经济研究领域进行大数据分析 & 提高预测效果的关键。在很多情况下,通过融合训练两种或两种以上有差异的算法,能够显著提高算法的泛化能力以提高预测效果。基于阿里巴巴电子商务平台购物行为数据,分别应用 Logistic 回归、支持向量机以及这两种算法的融合构建了预测模型。实证结果表明,融合后模型比单一模型具备更好的预测效果。

[13]葛绍林,叶剑,何明祥. 基于深度森林的用户购买行为预测模型[J]. 计算机科学, 2019, 46(09):190-194.

近年来,网络零售保持高速增长,网站中富含大量的用户行为数据。电商平台中的用户对商品的操作行为可以体现用户偏好,如何利用用户行为挖掘用户偏好已经成为学术界和工业界的关注焦点,并已经取得了众多研究成果。然而,目前用户操作行为预测方法研究通常只针对用户某一类操作行为进行分析,无法完备反映用户行为的整体特征。因此,提出一种基于深度森林的用户购买行为预测模型,通过构建用户行为特征工程建立整体用户行为特征模型;基于此,提出基于深度森林的用户购买行为预测方法,实现高效的行为预测训练效果。该方法的训练时间为 43 s,F1 值为 9.73%,相对其他模型取得了更好的效果。实验结果表明,该模型在降低时间开销的同时,提高了预测准确率。

[14]Huijuan H ,Dingju Z ,Tao W , et al.User Consumption Behavior Recognition Based on SMOTE and Improved AdaBoost[J]. International

Journal of Software Science and Computational Intelligence  
(IJSSCI), 2022, 14(1):1-20.

The sudden outbreak of COVID-19 has dealt a huge blow to traditional education and training companies. Institutions use the WeChat platform to attract users, but how to identify high-quality users has always been a difficult point for enterprises. In this paper, researchers proposed a classification algorithm based on SMOTE and the improved AdaBoost, which fuses feature information weights and sample weights to effectively solve the problems of overfitting and sample imbalance. To justify the study, it was compared with other traditional machine-learning algorithms. The accuracy and recall of the model increased by 19% and 36%, respectively, and the AUC value reached 0.98, indicating that the model could effectively identify the user's purchase intention. The proposed algorithm also ensures that it works well in spam identification and fraud detection. This research is of great significance for educational institutions to identify high-quality users of the WeChat platform and increase purchase conversion rate

[15]陈宁琦. 基于 LightGBM 算法的线上消费者购买意向研究[D]. 华中科技大学, 2020. DOI:10. 27157/d. cnki. ghzku. 2020. 006259.

随着互联网的普及,移动应用和电子支付技术的不断发展,电子商务突破了时间和空间的限制。消费者能随时随地消费的同时,也期待更优质的购物体验 and 个性化服务。现下常见的电商营销手段如发放优惠券和商品推荐都是基于消费者在线行为数据分析得到的。通过分析用户的历史行为数据,精准预测购物意向,能进一步为消费者提供有针对性的高质量服务,促使其消费进而提高购买转化率。因此如何有效利用消费者行为数据,分析其购物需求是所有电商企业面临的挑战。基于上述研究背景,本文旨在提出一个在线消费者行为分析系统,利用

消费者行为数据构建数据挖掘模型,综合不同的影响因素对消费者购买决策的影响,最终对用户的购买意向做出预测。根据不同的购物意向,可以提供个性化的服务,提高购买转化率。论文主要研究工作有:(1)消费者行为数据预处理和特征工程。重点研究了异常值处理、特征选择和不平衡数据过采样方法的原理以及实现。(2)构建模型。对支持向量机、随机森林、梯度决策树和 LightGBM 四种候选算法进行建模和调参,并对预测精度等四个方面进行综合评估。(3)对 LightGBM 模型的特征选择、基于代价敏感方法处理不平衡数据和划分购物意向三方面作进一步研究。实验结果表明,经过超参数优化的代价敏感 LightGBM 模型对电商消费者行为预测效果最优,AUC 值达到 0.893。综合评估发现该模型对消费者的购物意愿具有良好的预测精度、泛化能力以及实时性,一定程度上满足在线消费者行为预测系统的要求。此外还发现点击流数据传达了消费者购买意向的重要信息,利用分类器输出预测概率对用户的购物意向作进一步划分,能初步定位忠诚用户、潜在用户和流失用户,进而制定有针对性的营销策略,实现精准营销。本文通过以上研究,丰富了基于消费者行为数据对购买意向的理论基础,为在线消费者行为分析提供一种较好的模型,对建立实时在线消费者行为分析系统具有一定的理论意义和实践意义。

[16]张锐. 基于不平衡数据的保险购买预测研究[D]. 重庆大学, 2021. DOI:10. 27670/d. cnki. gcqdu. 2021. 000093.

随着技术日新月异的发展,“大数据”时代已经逐渐来临。对于保险公司来说,如果能够利用数据挖掘技术在海量的目标客户中寻找到高价值客户,将会对公司的业务发展产生重大意义。一般来说,在海量的目标客户中,有价值的客户远少于没有价值的客户,此时客户数据集是类不平衡的。那么如何从不平衡的目标客户中寻找出高价值客户对保险公司来说已经成为了重要的问题,解决好这个问题不仅可以转变保险公司传统的营销模式来提高业绩,而且还能促进保险行业的发展。本文主要研究基于不平衡数据的商业养老保险购买预测问题。首先,根据 2017 年的中国综合社会调查(CGSS)选取用于本文的研究数据,并对数据进行预处

理工作。以商业养老保险为研究对象,描述了商业养老保险在性别、年龄、个人年总收入、受教育程度、身体状况、户口等角度下的分布状况。利用假设检验的方法判断选取的影响因素与商业养老保险购买之间是否存在显著性差异,从而确定模型的输入变量。然后从抽样和算法两个方面处理不平衡问题,利用四种不同的方法建立模型,达到预测居民是否购买商业养老保险的效果。采用的方法分别为基于 SMOTE 方法的决策树模型、代价敏感决策树模型、代价敏感支持向量机模型以及 Adacost 模型。通过各类模型效果的比较,四种方法均较好地解决了不平衡问题,能够有效识别出购买保险的客户,并且在测试集上取得了较好的效果。其中 Adacost 模型的效果最为理想,具有实际应用价值。同时,根据代价敏感决策树模型和 Adacost 模型中的变量重要性大小,论文给出了影响居民购买商业养老保险的主要因素,对保险公司进行商业养老保险客户的识别具有指导作用。

[17]王予涵. 基于 LightGBM 的用户购买行为预测研究[D]. 兰州大学, 2022. DOI:10.27204/d.cnki.glzhu.2021.001505

电子商务的概念愈发火热,各种电商平台纷纷涌现,越来越多的人加入网购的大军,但当电商平台发展到一定程度后,流量的增加终究会停止,提高流量转化率无疑是一个重要且紧迫的课题.目前各电商平台都引入了推荐算法,为用户推荐其喜好的商品,提高用户体验,而预测是推荐的基础,提前预测出用户的购买倾向无疑会大大提高推荐算法的效果,这是一项极具意义的工作.基于此,本文选取京东算法大赛的数据来对用户购买行为预测进行研究,主要工作内容及成果如下:1.确定预测目标:在一段时间内有行为记录的用户-品类-店铺组合(称为 F<sub>i</sub>ID)中,预测未来 7 天会产生购买行为的 F<sub>i</sub>ID,这是预测中的二分类问题.2.确定训练集和预测集样本.选取 2018-03-19 到 2018-04-01 按照正负样本比为 1:30 负采样后共 517049 个有行为记录的 F<sub>i</sub>ID 作为训练样本,其中在未来 7 天发生购买的 F<sub>i</sub>ID 的标签为 1,其余为 0;选取 2018-03-26 到 2018-04-08 共 1792209 个有行为记录的 F<sub>i</sub>ID 作为预测样本,其中在未来

7 天发生购买的 F<sub>1</sub>ID 的标签为 1,其余为 0.3.构建基于时间滑动窗口的特征.本文从基本特征、累积特征、时间滑动窗口特征 3 个方面在用户、品类、店铺、用户-品类、用户-品类-店铺 5 个维度构建了 564 维特征,并对特征在缺失值等方面做了相应的处理.4.构建模型并选择最终的预测模型.本文利用 LR、RF、GBDT、XGBoost、LightGBM 在 517049\*565 的训练集上训练模型,并在 1792209\*565 的预测集上预测,从 AUC、F<sub>1</sub>分数、训练时间等方面比较分析,最终选取 LightGBM 作为最终的模型. .

下面开始全新的文献分析,分析利弊以及精度。下面需要你对这些篇参考文献进行总结,我会给你你所列出的第一条参考文献的例子,请你以此为模板将剩余的参考文献进行总结:

"肖励等人利用统计分析技术以及支持向量机、最小二乘支持向量机、XGBoost 以及 LightGBM,对银行客户购买产品意向数据进行了很好的数学建模,并详细介绍了其特征工程、模型训练以及模型对比分析,采用线性核函数的 LS-SVM 取得最好的效果,分类准确率达到 87.41%,但是对模型的能力挖掘不足以及对数据的正负样本不均衡挖掘不足。"

上面是例子请你分别对你所找到的剩余文献的"摘要部分"进行分析与汇总。直到输出完整才停止。每篇论文的精度是十分重要的,就像肖励论文中的分类准确率达到 87.31,每篇必须要有一个数据进行支持。

- [1] 肖励. 银行客户购买产品意向预测模型研究[D].中南财经政法大学,2020.
- [1] 谢青芳. 基于机器学习的用户购买行为预测 [D]. 兰州大学,2023.DOI:10.27204/d.cnki.glzhu.2023.003614.
- [2] 陈龙. 基于机器学习方法的用户复购行为预测 [D]. 南开大学,2021.DOI:10.27254/d.cnki.gnkau.2021.000097.
- [3] 梁宇婷. 基于过采样技术与机器学习的用户复购行为预测模型 [D]. 重庆大学,2021.DOI:10.27670/d.cnki.gcqdu.2021.001532.
- [4] 叶艳艳. 基于机器学习的电信用户流失预警模型预测与分析 [D]. 山东大学,2021.DOI:10.27272/d.cnki.gshdu.2021.003537.
- [5] 侯荣真. 基于不平衡数据集的机器学习算法对用户流失的预测 [D]. 南开大学,2021.DOI:10.27254/d.cnki.gnkau.2021.000070.
- [6] 胡雅竣. 基于机器学习的银行用户信用违约预测方法研究 [D]. 华中科技大学,2021.DOI:10.27157/d.cnki.ghzku.2021.002557.
- [7] 张晨楠. 基于机器学习的用户贷款风险预测模型研究 [D]. 海南大学,2020.DOI:10.27073/d.cnki.ghadu.2020.000268.
- [8] 张韶. 基于机器学习的用户购买行为预测研究 [D]. 长安大学,2020.DOI:10.26976/d.cnki.gchau.2020.000772.
- [9]Wenlong L .Default Prediction of Internet Finance Users Based on Imbalance-XGBoost[J].Tehnički vjesnik,2023,30(3):779-786.
- [10]Jing G ,Wenjun S ,Xin S .Research on Default Prediction for Credit Card Users Based on XGBoost-LSTM Model[J].Discrete Dynamics in Nature and Society,2021,2021
- [11] 常文晗. 基于 Adacost 算法的居民保险购买行为影响因素分析 [D]. 天津财经大



学,2020.DOI:10.27354/d.cnki.gtcjy.2020.000752.

[12]祝歆,刘潇蔓,陈树广等.基于机器学习融合算法的网络购买行为预测研究[J].统计与信息论坛,2017,32(12):94-100.

[13]葛绍林,叶剑,何明祥.基于深度森林的用户购买行为预测模型[J].计算机科学,2019,46(09):190-194.

[14]Huijuan H ,Dingju Z ,Tao W , et al.User Consumption Behavior Recognition Based on SMOTE and Improved AdaBoost[J].International Journal of Software Science and Computational Intelligence (IJSSCI),2022,14(1):1-20.

[15]陈宁琦.基于 LightGBM 算法的线上消费者购买意向研究[D].华中科技大学,2020.DOI:10.27157/d.cnki.ghzku.2020.006259.

[16]张锐.基于不平衡数据的保险购买预测研究[D].重庆大学,2021.DOI:10.27670/d.cnki.gcqdu.2021.000093.

[17]王予涵.基于 LightGBM 的用户购买行为预测研究[D].兰州大学,2022.DOI:10.27204/d.cnki.glzhu.2021.001505.