

# 基于机器学习对5G潜在客户分析与挖掘

洪晓晴, 潘 珈, 栾瑶瑶, 李 敏\*

青岛大学数学与统计学院, 山东 青岛

收稿日期: 2023年3月14日; 录用日期: 2023年4月14日; 发布日期: 2023年4月25日

## 摘 要

随着通信网络工程和新型基础设施技术的不断发展与完善, 我国正逐渐实现从4G社会向5G社会的转型。5G其低时延、大带宽、广连接的技术优势, 成为建设智慧城市和数字乡村重要的技术背景。为实现智慧城市建设所需的5G网络大规模连接条件, 需要更高的5G用户使用率。基于此问题, 本文从某移动大数据平台获取数据, 基于5G潜在用户的预测问题利用机器学习建立分类预测模型, 正确识别出潜在的5G用户并对其进行精准业务推荐, 提升我国5G使用率, 推进新型智慧城市建设的快速升级。构建预测模型的过程主要包括数据预处理、特征工程、模型的训练和评估。首先对数据进行预处理及探索性分析, 针对数据进行了包含数据清洗、去除唯一值属性、数据变换等在内的一系列预处理工作, 随后通过卡方检验、独立样本T检验和皮尔逊相关系数法对本文数据集中的特征进行了变量筛选, 筛选出特征重要度高的24个特征变量。根据筛选出的特征变量构建模型, 包括随机森林模型、CatBoost模型、LightGBM模型并进行参数调优, 寻找最优参数。根据得到的最优参数建立模型并通过测试集进行测试, 通过准确率、召回率、AUC值指标进行模型评价, 对比发现LightGBM模型对于5G潜在用户预测效果总体上优于其他模型。另外通过上述模型得到特征的重要性得分并进行重要性排序。通过本文方法实现对5G潜在用户较为准确的识别与挖掘, 运营商可据此实现对不同客户的精准营销, 推进更多用户实现4G向5G的转变, 加快我国5G市场的持续发展和智慧化城市的建设。

## 关键词

5G潜在客户识别, 随机森林, Catboost模型, LightGBM模型

# Analysis and Mining of 5G Potential Customers Based on Machine Learning

Xiaoqing Hong, Jia Pan, Yaoyao Luan, Min Li\*

School of Mathematics and Statistics, Qingdao University, Qingdao Shandong

Received: Mar. 14<sup>th</sup>, 2023; accepted: Apr. 14<sup>th</sup>, 2023; published: Apr. 25<sup>th</sup>, 2023

\*通讯作者。

## Abstract

With the continuous development and improvement of communication network engineering and new infrastructure technologies, China is gradually realizing the transition from a 4G society to a 5G society. 5G, with its technical advantages of low latency, large bandwidth and wide connectivity, has become an important technical background for the construction of smart cities and digital villages. In order to achieve the conditions for large-scale connectivity of 5G networks required for the construction of smart cities, a higher utilization rate of 5G users is required. Based on this problem, this paper obtains data from a mobile big data platform, builds a classification prediction model based on the prediction problem of potential 5G users, correctly identifies potential 5G users and makes accurate service recommendations to them, improves the 5G utilization rate in China, and promotes the rapid upgrade of the construction of new smart cities. The process of building the prediction model mainly includes data pre-processing, feature engineering, training and evaluation of the model. Firstly, data pre-processing and exploratory analysis were performed, and a series of pre-processing work including data cleaning, removal of unique value attributes, data transformation, etc. were carried out for the data, followed by variable screening of the features in the dataset of this paper through chi-square test, statistical t-test and Pearson correlation coefficient method, and 24 feature variables with high feature importance were screened out. Models were constructed based on the screened feature variables, including Random Forest model, CatBoost model, and LightGBM model, and parameter tuning was performed to find the optimal parameters. The models are built according to the obtained optimal parameters and tested by the test set, and the models are evaluated by accuracy, recall, and AUC value indexes, and the comparison reveals that the LightGBM model is generally better than other models for 5G potential user prediction. In addition, the importance scores of the features are obtained by the above model and ranked in importance. Through the method of this paper to achieve more accurate identification and mining of 5G potential users, operators can accordingly realize accurate marketing for different customers, promote more users to realize the transition from 4G to 5G, and accelerate the sustainable development of China's 5G market and the construction of smart cities.

## Keywords

5G Potential Customers Identification, Random Forest, Catboost Model, LightGBM Model

Copyright © 2023 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

伴随着移动智能设备的大规模普及，移动通信技术在人们的工作和生活中饰演了不同寻常的角色，第五代移动通信技术——5G (5th Generation Wireless Systems)是目前已投入使用的最为先进的通信技术[1]。近年来，社会转型加速，国家正在加强培育数据要素市场、推进治理体系现代化、推进新型基础设施建设，致力打造全新智慧城市。而5G网络的大规模连接能力、高速率传输能力正是智慧城市建设的有力支撑。

5G可高效将城市系统和服务打通、集成，提升资源运用效率，优化城市管理和服 务，改善市民生活质量。加快5G用户增长与城市发展深度融合，通过信息化手段解决城镇化过程中带来的问题，既是城市可持续发展所需，也是产业新动能所在。而如何通过模型精准识别5G需求潜在用户，促进4G时代向5G时代转变，以实现基于5G深度应用的智慧城市建设至关重要。

近些年来，数据挖掘技术快速发展，帮助人们更高效地解决了一些实际问题，机器学习等方法被广

泛应用于医学诊断、图像处理等各行各业。机器学习的诸多方法和手段，同样可以为竞争日益激烈的通信运营商领域挖掘潜在用户提供指导。本文选取行业代表性企业作为研究对象，选取了某市 5 万条数据进行二分类建模分析。针对其个人市场业务，基于它拥有大体量客户及海量客户信息的特点，探讨如何开展精准的客户分群营销，提前预警消费流失客户，做好挽留；同时充分挖掘有消费潜力的客户，拉升消费水平，对个人通信客户的消费能力进行深度挖掘，确保移动公司收入保持较高的增长速度。本论文构建 5G 潜在用户的识别模型，利用大数据优势建立模型，更精准地找到目标用户，可以有效地降低运营商营销成本，并且拓宽运行商行业发展，实现更多 4G 用户向 5G 的转型，提升我国的 5G 用户使用率和覆盖率。

## 2. 国内外研究现状

国内对于通信行业的发展还是走在世界前列的，通信行业随着网络技术和社会的发展，用户对运营商在各时期有着不同的要求。

2016 年，周意[2]利用大数据对江西移动 4G 营销情况做出分析，对 4G 终端客户的换机行为、用户画像、场景做出分析及营销策略。2020 年，董喆人[3]在 5G 背景下，探究目前流量增收为先的情况下对公司的流量营销做了精准营销优化的研究，为各大运营商推广流量营销给出一定的借鉴和参考意义。同年，欧阳秀平[4]利用机器学习构建和用户信息构建终端品牌推广模型。在识别潜在用户方面，郭林雪[5]借助关联规则对用户的网上购车行为特征进行分析，并根据此类用户的特点，利用协同过滤算法寻找出与老用户有相似行为特征的潜在用户。2022 年，周雅婷[6]首先采用 KNN 算法填补法等三种方法处理缺失值。在模型搭建前，对数据进行了描述性统计分析，可以系统地了解众多变量的分布特征。并搭建了 Logistic 回归模型、决策树分类模型、随机森林模型、XGBoost 模型和 LightGBM 模型五种模型，并将各个模型在准确率、召回率、AUC 这三个评估指标上的表现进行对比分析。

国外对于识别出潜在客户的研究也非常有成果，Kazemi 等人[7]通过分析顾客购物车中相关商品的信息，对顾客的购买行为进行挖掘，最后借助决策树这一算法对该公司的潜在客户进行识别。Coussement 等人[8]在邮件广告这种直接式营销策略中挖掘潜在用户，通过对客户与公司历史邮件链接之间的回复概率建模，从而识别出潜在客户。

## 3. 文章结构(图 1)

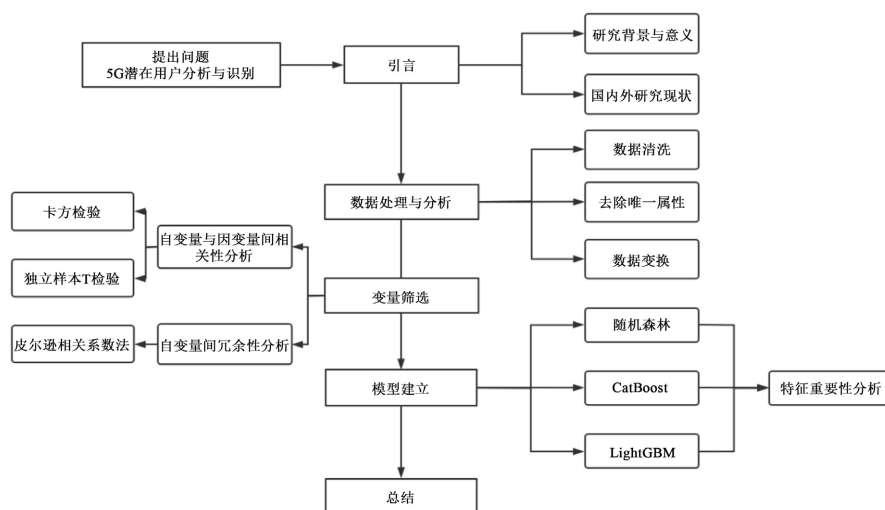


Figure 1. Article structure flow chart  
图 1. 文章结构流程图

## 4. 数据处理与分析

### 4.1. 数据来源与变量说明

本文通过爬取某省移动大数据平台的信息得到原始数据 50,000 条, 包括 44 个变量。由于变量较多, 为了更好的做可视化处理, 我们将 44 个变量建立 10 个维度, 分别为用户标识、用户基础信息、消费行为信息、超套信息、宽带信息、签约信息、套餐信息、流量饱和度信息、其他信息、标签。

### 4.2. 数据预处理

在数据挖掘过程中, 海量的原始数据中往往存在大量的异常数据, 这些数据的存在会使得建模的执行效率降低, 造成结果的偏差。为了提高数据的质量, 得到更好的建模效果, 需对原始数据预处理。通过对本文中的数据集进行初步探索, 本次操作对数据进行的操作包括数据清洗、去除唯一属性、数据变换等。

#### 4.2.1. 数据清洗

数据清理是一道能够发现并纠正数据中可辨别错误的程序, 从而提高后续所研究数据集的质量, 本研究过程中主要进行的操作包括对缺失和重复数据的处理、一致性检查等。

##### 1) 缺失值处理

缺失值是指当前数据集中某个案的某个或某些特征信息是缺失的。在处理缺失值时, 按分析缺失来源分可以考虑从个案或特征两个维度处理, 按缺失数据处理手段分为删除缺失值和插补缺失值。本文主要基于统计学原理填充缺失值, 从而使信息表完备化。本文有两类缺失值, 一类为名义型变量缺失值, 另一类归为连续型变量缺失值。针对名义型变量缺失值, 我们使用该类别的众数进行填充。针对连续型变量缺失值, 我们利用其余对象在该特征上取值的平均值来插补缺失值(表 1)。

**Table 1.** Missing variables and quantity  
**表 1.** 缺失变量及数量

变量名称	缺失数量	变量名称	缺失数量
X3	2161	X20	133
X5	217	X21	133
X6	36	X22	133
X7	36	X23	133
X8	36	X26	32,329
X9	36	X27	32,329
X10	36	X28	241
X11	36	X29	241
X12	36	X30	241
X13	36	X31	241
X14	36	X32	2083
X15	36	X33	2083
X16	36	X34	2916
X17	36	X35	3149
X18	133	X36	3072
X19	133	X38	42,608

## 2) 重复值处理

重复值是指数据集中出现属性值完全相同的数据，即某一条或者多条用户数据重复出现。根据用户编码这一唯一标志对数据集中的数据进行检索，发现其中重复值较少，删除重复数据对整体数据影响不大，故对重复值做去除处理。

## 3) 一致性检查

一致性检查将变量合理的取值范围作为依据，观察数据是否合乎实际要求，及时地发现超出合理范围的数据。如本研究对用户的性别、年龄等属性进行探索，性别为男女两种类型，年龄范围在 18~60，均属于正常范围。

### 4.2.2. 去除唯一属性

唯一属性是指唯一标识实体实例的属性，通常是指一些 ID 属性，这些属性并不能刻画样本本身的分布规律，一般不会用在建模之中，可直接予以删除。例如本研究中的 `user_id` 和 `product_no` 是随机生成的标志用户的属性，并不具有业务意义，故进行删除处理。

### 4.2.3. 数据变换

数据变换是指对原始数据中的字段数据进行一定的转换操作，从而得到适合进行算法模型构建的输入数据(数值型)，以便于后续的信息挖掘。在本研究数据中，X1 (性别)和 X5 (细分市场)这两个属性为文本型数据，不符合机器学习模型对于数值型输入变量类型这一要求，需要进行文本数据转换，也就是将文本数据转换为数值型的数据。在对文本类型的该属性数据进行转换时，考虑到频次和重要性双方面的因素，我们发现 X1 只有两个类别：女士、先生；X5 也是两个类别：农村用户、大众用户、校园用户、集团用户。两个属性的变换形式如下表 2 所示：

Table 2. Text data conversion

表 2. 文本型数据转换

属性	数据转换
X1	女士：0；先生：1
X5	农村用户：0；大众用户：1 校园用户：2；集团用户：3

## 5. 变量筛选

### 5.1. 自变量与响应变量间相关性分析

不同类型数据在探究自变量和响应变量之间相关性有不同的方法。针对分类自变量，可以通过卡方检验探究自变量与响应变量之间相关性；针对连续性变量，使用独立样本 T 检验来分析其与响应变量之间的关系。综上，通过分别使用上述两种方法进行变量的初步筛选，根据与评价指标进行比较舍弃无关变量。

### 5.2. 皮尔逊相关系数法

根据上述步骤得到的自变量对目标变量有较为显著的影响，但并没有对自变量之间相关性进行探寻，若自变量之间有较大的相关度，容易造成较大冗余性，会对 5G 潜在用户的识别造成误差。因此本节需要对自变量之间的相关性进行分析。

通常，两变量 X 和 Y 间的相关性由皮尔逊相关系数  $r$  来表示，根据  $r$  的取值大小判断变量之间的相

相关性, 当  $|r|$  大于 0.6 时, 认为两个变量之间达到了较强的相关性, 对该类变量进行进一步的筛选。其计算的公式为:

$$r = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2} \sqrt{\sum_{i=1}^n (Y_i - \bar{Y})^2}} \quad (1)$$

利用上式计算各个自变量的相关系数  $r$  并绘制各变量的相关性热力图如下图 2 所示, 图中颜色越接近红色说明两者之间的相关性较强, 不难发现某些变量之间的相关性达到 0.6 以上甚至达到 0.9。结合各指标的背景和定义, 剔除部分冗余变量, 例如变量近三月平均 mou, 与当月 mou、上月 mou、上上月 mou 之间的相关性达到 0.9 以上, 若将其同时纳入模型中, 会造成最终预测的较大误差。根据该变量的定义(近三月每户的平均通话时长), 该变量可以看作作为扩展变量, 是对后三个变量的概括总结, 因此可以删除后三个变量。

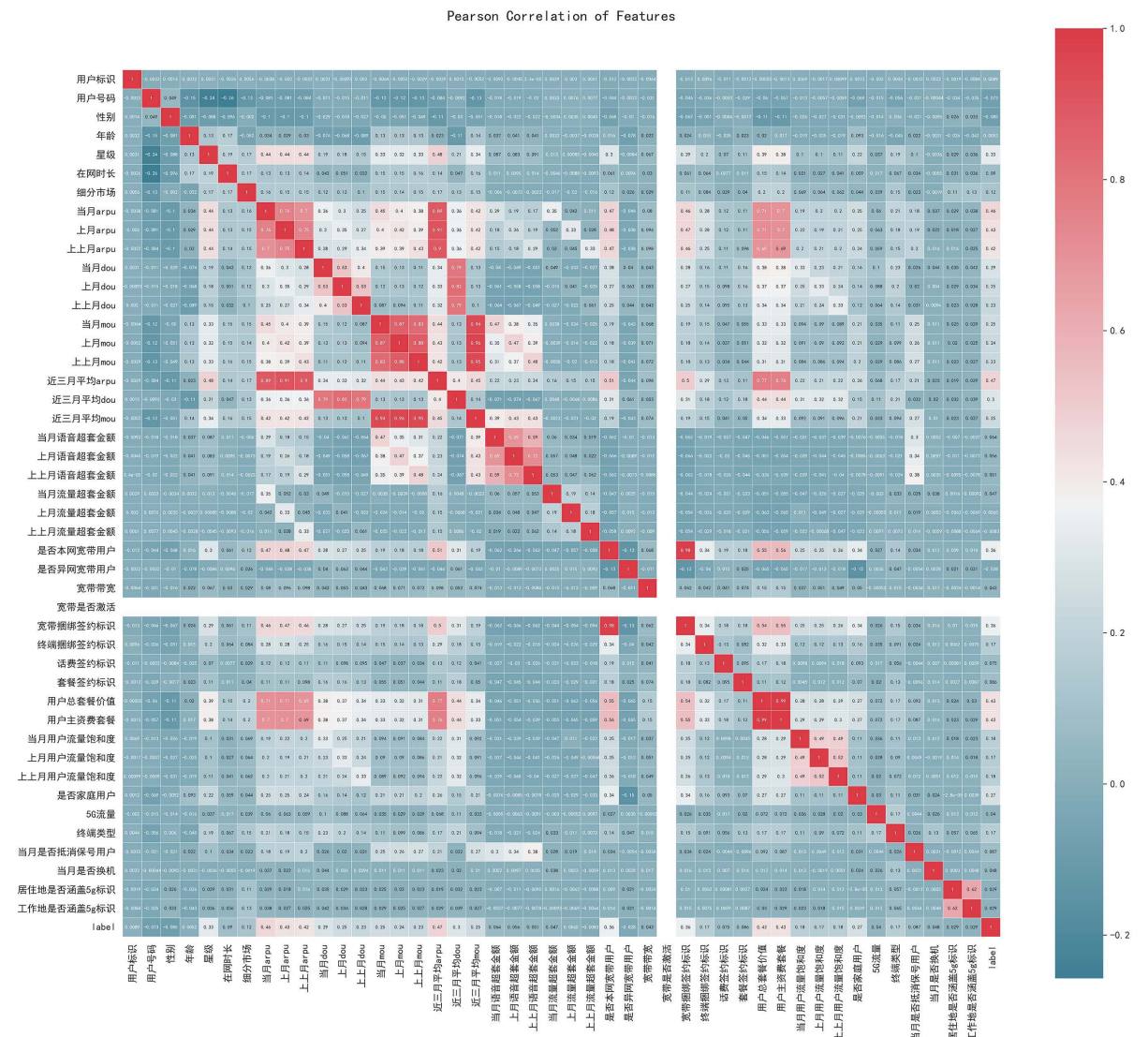


Figure 2. Correlation heat map  
图 2. 相关性热力图

## 6. 模型预测

### 6.1. 模型评价指标

建立模型之后，需要根据合适的评价指标对模型效果进行评价，选择更优的模型。本文的任务是对 5G 目标客户的分类，常见的分类评价标准包含准确率、召回率、AUC 值。二分类混淆矩阵实例是分类指标的基础，如下表 3 所示：

**Table 3.** Obfuscation matrix example  
**表 3.** 混淆矩阵实例

	正样本预测用户	负样本预测用户
正样本实际用户	A	C
负样本实际用户	B	D

1) 准确率(Accuracy)是所有类别用户预测正确的比率。

$$\text{准确率} = \frac{A + D}{A + B + C + D} \times 100\% \quad (2)$$

2) 召回率(Recall)是真正样本用户中被预测结果为正样本用户的比例，通过该评价指标能够反映少数样本类别(开通 5G)被正确预测的比例。本文目标变量的各类别数量相差较大，能否正确识别少数样本类别(开通 5G)很重要，因此需要考虑该指标的值。

$$\text{召回率} = \frac{D}{B + D} \times 100\% \quad (3)$$

3) AUC 值：ROC 曲线下半部分的面积，其范围区间在[0, 1]之间，面积越大则模型效果越好，此时曲线尽可能的处在左上角，这就能证明模型在预测正确的概率大于模型预测错误的概率。

### 6.2. 随机森林

随机森林是一种由决策树构成的集成算法，属于集成学习中的 Bagging 方法。由于单一决策树可能会造成过拟合和较大的误差，该算法尝试将多个不同的决策树结合起来，从而减少单一决策树可能存在的缺陷和判断不准确的问题。本文的任务是对 5G 潜在用户的识别，属于分类问题。当输入新的样本时，森林中的每一个决策树对其进行判别与分类，对每一颗树的分类结果进行归类汇总，将分类结果最多的那一类作为随机森林的最终结果。

在 Python 软件中利用 RandomForestClassifier 函数可以实现随机森林的分类问题，其主要参数包括：决策树的棵数(n\_estimators)、构建决策树最优模型时考虑的最大特征数(max\_features)、叶子节点最少样本数(min\_samples\_leaf)、最小样本划分的数目(min\_samples\_split)等。合理设置参数可以提高模型的预测效果。本文使用网格搜索法 GridSearchCV 函数进行调参，将梯度优化结果的最佳参数作为网格搜索的最终参数，其中最优参数如表 4 所示。将最优参数再次带入模型预测，得到评价指标结果如表 5 所示。

### 6.3. CatBoost

Catboost 算法是一种带有分类特征的无偏增强算法，该算法引入处理分类特征的创新算法以及有序提升方法这两类方法，克服了在当前现有的梯度增强算法，普遍存在的后续梯度增强依赖训练集所产生的梯度偏差问题。

**Table 4.** Optimal parameter values for random forest**表 4.** 随机森林最优参数取值

参数名称	取值
criterion	entropy
n_neighbors	130
min_samples_split	46
min_samples_leaf	4
max_features	auto

**Table 5.** Random forest evaluation index results**表 5.** 随机森林评价指标结果

评价指标	百分率
准确率	0.8768
召回率	0.8610
AUC	0.8210

#### 1) 分类特征目标统计

传统梯度提升算法，经常将离散变量类别转化为数值型处理，处理时按照特征高低分成两类进行处理，低维特征时可以利用 One-hot，当遇到高纬度特征时 One-hot 转化会造成维度灾难的问题。为了解决这个问题，提出分类特征目标统计方法，对统计的目标变量进行分组后，用一个新的变量代替分组有效减少分类维度方法。

#### 2) 有序提升方法

预测偏移是由梯度偏差影响造成的，具体表现为在梯度增加的迭代时，训练时梯度偏差误估，导致模型过拟合。有序提升一种增强算法，它不会受到预测偏移中梯度偏差的影响，具体方式表现为在梯度提升的每一步，独立地采样一个新的数据集，通过将当前模型应用于新的训练示例，获得未偏移的残差。

同样对模型进行参数调优，选取的最优参数结果如下表 6。接下来，根据得到的最优参数在测试集上建立模型并进行评估，得到预测效果如下表 7 所示：

**Table 6.** Optimal parameter values for Catboost**表 6.** Catboost 最优参数取值

参数名称	参数值
iterations	1000
学习率	0.05
max_depth	7

**Table 7.** Catboost evaluation metrics results**表 7.** Catboost 评价指标结果

评价指标	百分率
准确率	0.9198
召回率	0.8646
AUC	0.8987



## 6.4. LightGBM

LightGBM (Light Gradient Boosting Machine)是梯度提升算法的一种，能够高效处理海量数据。该算法效率达到如此之快，是基于改进的减少样本角度和减少特征角度两类方法，单边梯度采样 GOSS，通过保留较大梯度的样本，排除大量的小梯度样本，从而减少样本量；互斥特征绑定 EFB，当样本特征较多时，将非零特征排序计算比率，使用合并特征获得最小比率。这两种方法克服了较流行的 GBDT 利用二阶导数进行梯度增强，在特征维数较大和数据量较大时，效率和可伸缩性稍差等缺点。

LightGBM 算法同 XGBoost 类似，均属于 Boosting 集成算法中的一种，即在前一轮预测结果基础上进行改进，但是二者的优化方式不同，XGBoost 是按层分裂，进行预排序分割，LightGBM 算法则按叶子节点的分裂方式进行决策树分裂，并基于直方图算法将特征值分桶再分割，相比前者而言其运行速度得以提升，内存资源占用更低，能够更快速地处理海量数据。本文选取的样本数量较多，为 50,000 条用户信息，因此将 LightGBM 算法应用于潜在用户预测研究中，充分发挥该算法灵活可靠的特点。

LightGBM 参数设置较为复杂，同样需要对重要参数进行调整，以提升模型的分类能力。通过网格调优搜索算法最终设置的主要参数结果如表 8 所示。将上述根据网格搜索出来的最佳参数值带入模型进行运行，并将模型带入测试集中进行预测效果分析，得到如表 9 预测效果。

**Table 8.** Optimal parameter values for LightGBM

**表 8.** LightGBM 最优参数取值

参数	参数值
提升算法	gbdt
目标函数	二分类
分割标准	Gini
学习率	0.1
弱学习器的最大迭代次数	100
树的最大深度	6
叶子节点上所需的最小样本权重	1
每棵树上拥有的最大叶子数	16
子特征处理列采样数	0.8

**Table 9.** LightGBM evaluation metrics results

**表 9.** LightGBM 评价指标结果

评价指标	百分率
准确率	0.9103
召回率	0.8703
AUC	0.8835

由上表结果可以看出，lightGBM 模型预测效果良好，预测准确度达到了 90%。也就是说，利用该模型进行预测，可以较准确的分析出 5G 潜在客户，可以更精准的定位顾客。

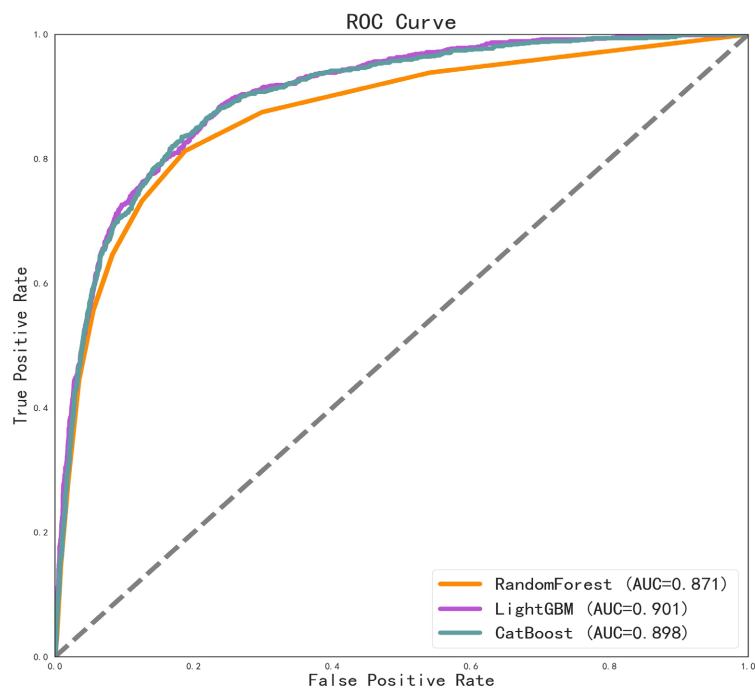
## 6.5. 模型比较

上述分别用训练集训练对应的潜在用户预测模型，并用测试集来评估模型的潜在用户预测效果。通

过网格搜索法选择最优参数从而提高了模型的预测准确度。模型评价指标是检验一个模型训练效果的重要标准，我们用准确率、召回率、AUC 三个指标对模型效果进行评定，经训练后的各个模型指标如表 10 所示。绘制各模型的 ROC 曲线如图 3 所示。

**Table 10.** Comparison of different model evaluation indicators for 5G potential users  
**表 10.** 5G 潜在用户不同模型评价指标比较

模型	准确率	召回率	AUC 值
随机森林	0.8768	0.8660	0.8652
CatBoost	0.9198	0.8646	0.8987
LightGBM	0.9103	0.8713	0.9031



**Figure 3.** ROC curves for each model

**图 3.** 各模型 ROC 曲线图

由表 10 可知，基于 Boosting 的 CatBoost 模型和 LightGBM 模型的各项评估指标均优于基于 Bagging 的随机森林。通过 ROC 曲线图可以看出 LightGBM 模型的召回率值较高，说明该模型能够较好的捕捉少数类。且 LightGBM 和 CatBoost 模型的准确率也略高，说明该类模型的鲁棒性较强，在捕捉少数类的同时也保障了多数类的平衡。LightGBM 和 CatBoost 模型在预测性能的表现上相差不大，但 CatBoost 模型存在缺点，即对于类别型特征的处理需要大量的内存和时间。因此，通过对以上模型在各个指标上的表现效果分析可知，各模型表现效果按照优劣进行综合排序：LightGBM 模型、CatBoost 模型、随机森林模型。

## 6.6. 特征重要性分析

本文的目的是对 5G 潜在用户的识别，同时寻找最能影响用户是否会选择 5G 的重要特征，根据这些

重要因素实现运营商对客户的初步划分，对不同的客户制定不同的营销策略。本文建立的模型均可以得到特征的重要性得分，对重要性得分前 10 名的特征进行重要性排序，得到下图 4。将图中三个模型同时存在的特征确定为重要特征，其中包括：近三月 arpu、是否为家庭用户、用户总套餐价值等。说明这些特征在构建 5G 潜在用户识别模型中发挥重要作用，在今后的研究分析中应加以重视。

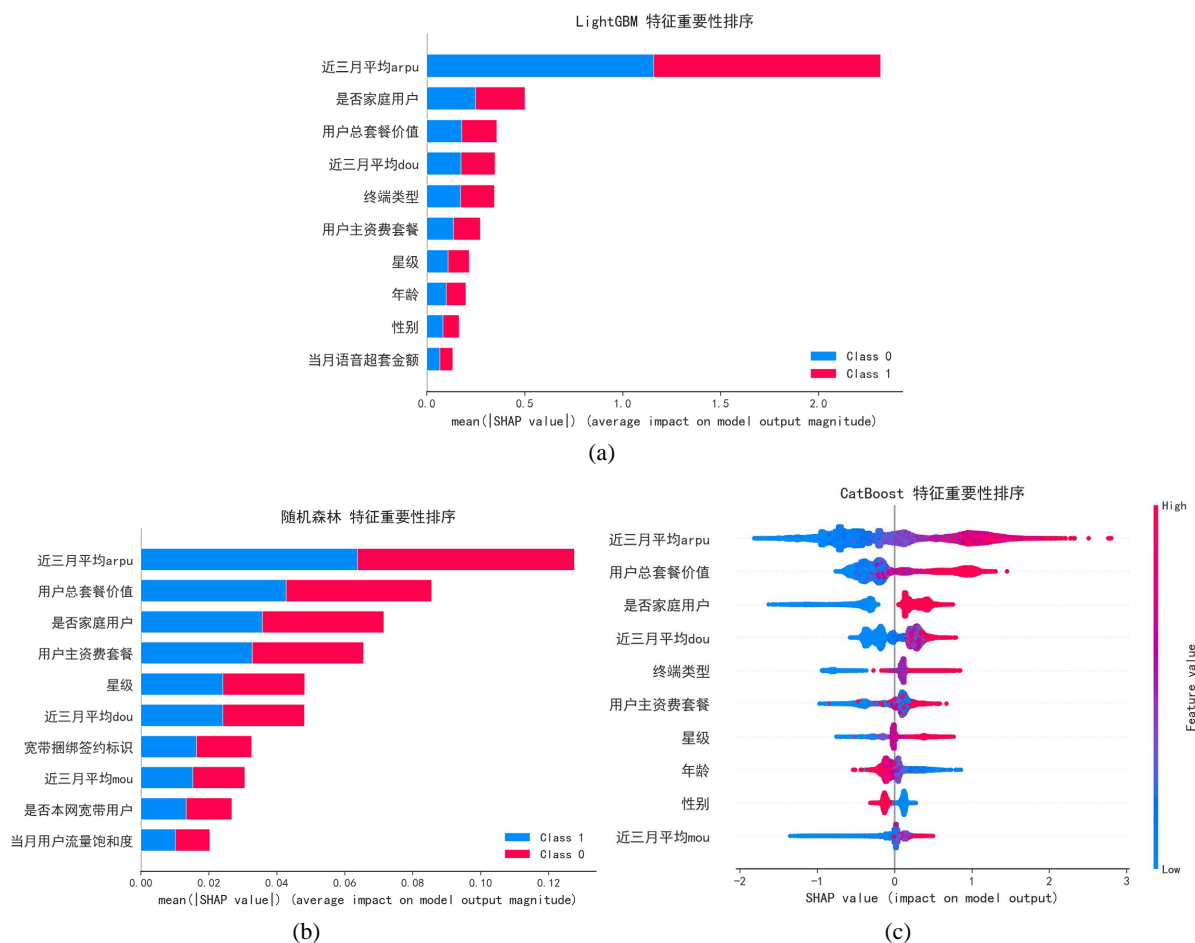


Figure 4. Ranking the importance of each model feature

图 4. 各模型特征重要性排序

## 7. 总结

本文利用某省移动 5G 用户大数据，挖掘理解用户需求，根据现有 5G 用户的使用行为信息建立潜在用户数据预测模型，结合 5G 潜在用户数据的特点，以数据挖掘的流程为主线，在分类算法层面实现了对 5G 潜在用户的预测。

本文首先对原始数据进行了数据预处理，其中包括缺失值处理、重复值处理、去除唯一属性、数据变换等步骤，并对处理后的数据进行变量筛选，得到 24 个变量用于后续建模。针对现有的 5G 潜在用户建立分类预测模型，并考虑样本数据数量多的特点，本文选取的模型有常用机器学习算法：随机森林模型，集成学习算法：CatBoost 模型、LightGBM 模型，并利用网格搜索法进行模型调参以选取最优参数。结合准确率、召回率、AUC 值等指标对各模型的预测效果进行比较发现，CatBoost 模型以及 lightGBM 模型在各方面表现较好。

5G是新一轮科技和产业革命中的核心关键技术之一,截至2022年底,我国5G基站数超过231万个,但是5G的用户使用率仍然偏低,需尽快提升我国的5G用户率。本文通过统计方法与建模正确识别5G潜在用户,为运营商提供参考,对潜在的5G用户提供精准产品推销,实现更多用户从4G向5G的转型,推进我国智慧化城市和数字经济的发展。

## 基金项目

山东省自然科学基金青年项目(ZR2021QA053)。

## 参考文献

- [1] 熊文德,李华丽. 5G用户数扩散预测研究[J]. 科技与创新, 2019(18): 122-123.
- [2] 周意. 基于大数据江西移动4G终端营销策略研究[D]: [硕士学位论文]. 南昌: 江西财经大学, 2016.
- [3] 董喆人. 基于5G背景的YZ移动公司流量营销策略优化研究[D]: [硕士学位论文]. 南京: 南京邮电大学, 2020.
- [4] 欧阳秀平, 万源沅, 邹俊德. 基于机器学习的终端换机预测模型[J]. 邮电设计技术, 2020(4): 75-79.
- [5] 郭林雪. 关联规则及协同过滤在汽车电子商务中的应用[J]. 科技经济导刊, 2017(8): 31.
- [6] 周雅婷. 基于数据挖掘识别移动5G潜在客户[D]: [硕士学位论文]. 重庆: 西南大学, 2022.
- [7] Kazemi, A., Babaei, M.E. and Javad, M.O.M. (2015) A Data Mining Approach for Turning Potential Customers into Real Ones in Basket Purchase Analysis. *International Journal of Business Information Systems*, **19**, 139-158. <https://doi.org/10.1504/IJBIS.2015.069427>
- [8] Coussement, K., Harrigan, P. and Benoit, D.F. (2015) Improving Direct Mail Targeting through Customer Response Modeling. *Expert Systems with Applications*, **42**, 8403-8412. <https://doi.org/10.1016/j.eswa.2015.06.054>