

基于机器学习的航空乘客满意度预测

王 晨, 罗鄂湘*

上海理工大学管理学院, 上海

收稿日期: 2025年4月28日; 录用日期: 2025年5月21日; 发布日期: 2025年5月29日

摘 要

应用Transformer以及其他机器学习算法对航空乘客满意度进行综合评价, 分析航空乘客对航空服务质量的满意度及其影响因素并对其满意度水平进行预测研究, 为提高航空服务质量提供参考。首先通过文献调研发现, 大多数研究集中在2015年美国航空乘客满意度数据集上。为避免与现有研究重复, 本研究从kaggle平台选取了一个与2015年美国数据集特征相同且样本量较大的数据集, 通过合并这两个数据集, 进行一项更深入的分析。然后通过现有特征构建两两特征间相加和相乘的新特征, 通过用一个较为简单的模型LightGBM来对比看哪种特征处理方式获得的新特征最佳, 最后对比多个机器学习模型, 确定最优模型。结果显示, Transformer在准确率、召回率的三个指标均优于其他算法, 其中Accuracy: 0.9592, AUC: 0.9604, Precision: 0.9573, Recall: 0.9604, F1-score: 0.9586。在线登记是航空服务质量满意度的重要影响因素, 航空公司应进一步加强以上方面的培训建设。

关键词

航空服务, 满意度, Transformer, 机器学习

Machine Learning-Based Prediction of Air Passenger Satisfaction

Chen Wang, Exiang Luo*

Business School, University of Shanghai for Science and Technology, Shanghai

Received: Apr. 28th, 2025; accepted: May 21st, 2025; published: May 29th, 2025

Abstract

Transformer and other machine learning algorithms are used to comprehensively evaluate the satisfaction of air passengers, analyze the satisfaction of air passengers with aviation service quality

*通讯作者。

and its influencing factors, and predict the level of air satisfaction, so as to provide reference for improving the quality of aviation services. First of all, through literature research, it is found that most of the research focuses on the 2015 air passenger satisfaction dataset in the United States. In order to avoid duplication with existing studies, this study selected a dataset with the same characteristics as the 2015 US dataset and large sample size from the Kaggle platform and combined the two datasets for a more in-depth analysis. Then, through the existing features, a new feature is constructed to add and multiply between two features, and a simple model, LightGBM, is used to compare which new feature processing method is the best. Finally, multiple machine learning models are compared to determine the optimal model. The results show that Transformer is better than other algorithms in terms of accuracy and recall, Accuracy: 0.9592, AUC: 0.9604, Precision: 0.9573, Recall: 0.9604, and F1-score: 0.9586. In the decision-making of satisfaction. Online boarding is an important factor in aviation service quality satisfaction, and airlines should further strengthen the training construction in the above aspects.

Keywords

Aviation Service, Satisfaction, Transformer, Machine Learning

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. 引言

随着全球经济复苏, 国际交流合作推动航空业重回增长轨道。Aviation Week 数据显示, 2019~2024 年飞行时数呈现明显复苏趋势, 2024 年上半年全球客运飞行时数创历史新高。我国民航市场同步回暖, 航班量、客流量持续攀升, 机票搜索量倍增。2024 年 3 月, 多部门联合印发《通用航空装备创新应用实施方案(2024~2030 年)》, 提出到 2027 年基本建立现代化通航体系。但高铁网络的完善对民航中短途运输形成竞争压力。

民航业作为服务型行业, 乘客满意度直接关乎企业收入、品牌价值和市场竞争力。在出行需求日益多元化的背景下, 乘客感知价值和服务质量成为影响满意度的关键因素。然而, 航空满意度研究面临两大挑战: 一是调查问卷设计和样本多样性的影响, 二是民航数据体量大、复杂度高对分析算法提出的要求。

当前研究方法主要分为实证分析和机器学习两类。实证研究通过统计方法验证变量关系, 但存在数据不全、理论局限等缺陷。机器学习能处理海量数据并挖掘非线性关系, 更适合发现传统方法难以识别的复杂模式。方法选择需根据研究目标、数据条件等因素决定。

提升航空满意度需要精准识别关键影响因素。除服务质量、航班准点率、客舱环境等传统指标外, 乘客个性化需求和数字化体验等新兴维度也日益重要。航空公司应通过优化服务流程、提升员工素质、加强技术创新等方式全面提升乘客体验, 在行业复苏中把握发展主动权。未来研究可结合多源数据, 采用混合方法深入挖掘满意度驱动机制, 为民航服务升级提供更精准的决策支持。

结构方程模型(Structural Equation Modeling)是一种融合了因素分析和路径分析的多元统计技术, 它的优势在于对多变量间交互关系的定量研究。一直以来, SEM 大量应用于社会科学及行为科学领域。通过查阅文献发现结构方程模型(SEM)是航空乘客满意度研究中应用最为广泛的一种方法, 其目的在于探索事物间的因果关系, 并将这种关系用因果模型、路径图等形式加以表述。毛曼等(2010)借鉴顾客满

意度指数模型 ACSI、ECSI, 选用 SEM 方法对上海虹桥机场进行实证分析, 创新性地增加了“感知服务质量”这一新的变量, 使得模型更具实用性和针对性。通过定量分析各变量得出了顾客“感知服务质量”对满意影响最大[1]。Park 等(2019)运用结构方程建模方法研究了乘客对航空公司感知服务满意度的潜在决定因素, 结果发现顾客的情感价值对航空公司服务满意度有显著影响[2]。Huang (2023)同样利用结构方程建模, 调查了低成本航空公司的乘客满意度水平和影响整体服务体验的因素, 研究揭示了感知价值的调节作用, 这是之前研究中较少探讨的领域[3]。Shin 等(2024)利用结构方程模型(SEM)评估旁观者评价与分享意图之间的关系, 创新性地以乘客作为旁观者的角度出发, 分析飞行中断行为对整体服务质量的影响[4]。

利用机器学习进行满意度预测的研究也逐渐丰富。Gao 等(2021)不仅利用机器学习方法研究发现了服务属性对航空满意度的线性影响, 其对比研究还发现, 与传统的逻辑回归方法相比, 多层感知机、随机森林和支持向量机等机器学习算法在乘客满意度研究中表现出优越的预测性能, 随机森林以 95.92% 的极高准确率居首[5]。Park 等(2022)利用深度学习对韩国航空乘客满意度进行了研究, 评估了包括 KNN、CART、RF、极端梯度提升(XGBoost)以及 CNN 和 CNN 长短期记忆网络(CNN-LSTM)在内的多种模型, 以预测航空乘客满意度。结果显示, 深度学习模型在大多数情况下优于机器学习模型, 其中 CNN-LSTM 模型在预测满意度方面的准确度达到了 90% [6]。Pranav 等(2022)用逻辑回归、随机森林和决策树分类器作为第 1 层, XGBoost 分类器作为第 2 层的组合分类器的堆叠集成模型来预测航空客户满意度。对比最佳基础学习器随机森林, 集成学习堆叠模型准确率达到 96.26% [7]。Noviantoro 等(2022)运用了八种监督式机器学习算法: 决策树、随机森林、GBT、KNN、朴素贝叶斯、规则归纳、逻辑回归、神经网络、深度学习和支持向量机方法, 开发乘客满意度预测模型。结果表明, 深度学习准确率最高为 95.42% [8]。Jiang 等(2022)则提出了一种基于随机森林递归特征消除(RF-RFE)与逻辑回归(LR)的模型, 用以预测和分析航空乘客的满意度。该研究为航空公司提供了在 COVID-19 疫情严重挑战下, 如何通过改善服务来增强乘客满意度的参考, 模型准确度达到了 96.3% [9]。Song 等(2024)研究表明提出的预测模型在准确度上超越了传统的机器学习基准模型, 基于广义线性模型(GLM)和深度神经网络(DNN)结合的深度学习模型进行满意度预测, 其准确率达到 92.9% [10]。

以上文献都为理解航空乘客满意度影响因素和预测航空乘客满意度方面提供了多维度的视角和深入的见解, 并且研究人员能够为航空公司提供关于如何改善服务和增强乘客满意度的深入建议。

本文创新性地建立了基于 transformer 算法的航空乘客满意度分类预测模型, 筛选出满意度高的影响因素, 帮助航空公司管理层了解乘客如何看待他们的航空公司服务, 更好地使用客户驱动的服务评估方法。本研究可为航空公司提供适当的服务水平以回应顾客需求提供指导。

2. 理论基础

2.1. 算法介绍

逻辑回归模型的基本思想是使用逻辑函数(Sigmoid 函数)将线性组合的输入映射到 0 和 1 之间, 从而实现对样本属于某一类别概率的预测。对于给定的输入特征向量 x , 逻辑回归模型的输出为式(1):

$$h_{\theta}(x) = \frac{1}{1 + e^{-\theta^T x}} \quad (1)$$

其中, θ 是模型的参数向量, T 表示转置。

L1 正则化, 也称为 Lasso (Least Absolute Shrinkage and Selection Operator), 通过在损失函数中添加参数的绝对值之和来约束模型的复杂度。L1 正则化项的形式为式(2):

$$\|\theta\|_1 = \sum_{i=1}^n |\theta_i| \quad (2)$$

其中, n 是特征的数量, θ_i 是参数向量 θ 的第 i 个元素。

对于逻辑回归模型, 常用的损失函数是交叉熵损失(Log Loss), 其形式为式(3):

$$J(\theta) = -\frac{1}{m} \sum_{i=1}^m \left[y^{(i)} \log(h_\theta(x^{(i)})) + (1 - y^{(i)}) \log(1 - h_\theta(x^{(i)})) \right] \quad (3)$$

其中, m 是训练样本的数量, $y^{(i)}$ 是第 i 个样本的真实标签(0 或 1), $x^{(i)}$ 是第 i 个样本的特征向量。在 L1 正则化约束下, 损失函数需要加上 L1 正则化项, 得到式(4):

$$J(\theta) = -\frac{1}{m} \sum_{i=1}^m \left[y^{(i)} \log(h_\theta(x^{(i)})) + (1 - y^{(i)}) \log(1 - h_\theta(x^{(i)})) \right] + \lambda \|\theta\|_1 \quad (4)$$

其中, λ 是正则化系数, 控制正则化项的权重。较大的 λ 值会增加正则化效果, 使得模型更加倾向于参数的稀疏性, 从而实现特征选择。

GBDT (Gradient Boosting Decision Tree) 的主要思想是利用决策树迭代训练以得到最优模型, 该模型具有训练效果好、不易过拟合等优点。但是面对海量的数据, 普通的 GBDT 算法是不能满足其需求的, LightGBM (Light Gradient Boosting Machine) 的提出可以更好更快地用于大规模数据。

LightGBM 的提出是基于决策树算法的梯度提升框架。它具有高效、准确和高可扩展性的特点, 能够在海量数据上快速训练模型, 并在不牺牲精度的前提下提高处理速度。

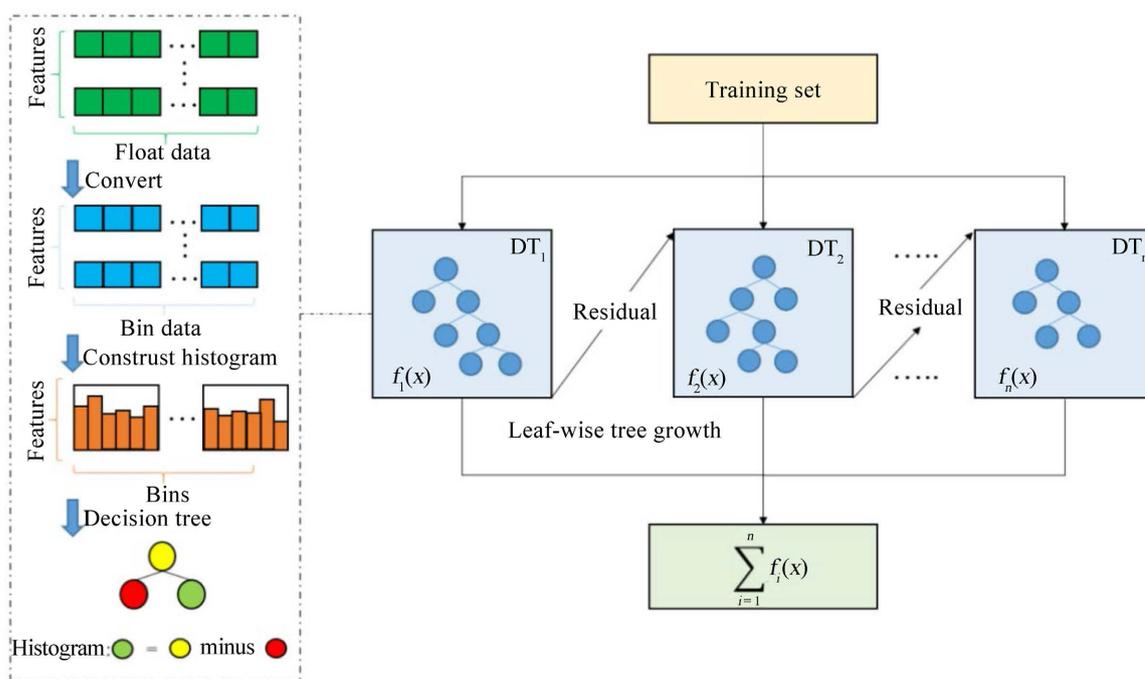


Figure 1. LightGBM schematic
图 1. LightGBM 原理图

LightGBM 相较传统的 GBDT 框架进行了多项改进, 如图 1 所示, 主要包括:

- 1) 基于叶子节点的增长策略: 在每次迭代中选取所有叶子节点中增益最大的节点进行分裂, 从而生成一个非对称的树结构。
- 2) 基于直方图的算法: 将连续的特征离散化为有限的直方图桶, 加快了计算速度并降低内存占用。

- 3) 基于特征的单边梯度采样: 保留较大梯度的样本, 随机采样较小梯度的样本, 加速模型训练。
- 4) 独特的特征互斥机制: 将稀疏特征进行捆绑处理, 进一步减少特征维度。

2.2. 技术路线

首先对数据进行预处理, 包括缺失值处理、删除重复值等。然后进行特征处理, 涵盖特征缩放、特征编码、特征构建等步骤。在模型构建阶段应用多种机器学习算法, 如决策树、逻辑回归、随机森林、K 近邻等。最终选择最优模型并展示结果。整个流程系统化地涵盖了从数据预处理到模型评估的各个环节, 为航空乘客行为分析提供了全面的技术框架。

图 2 是运用各类算法进行满意度预测的技术路线图。

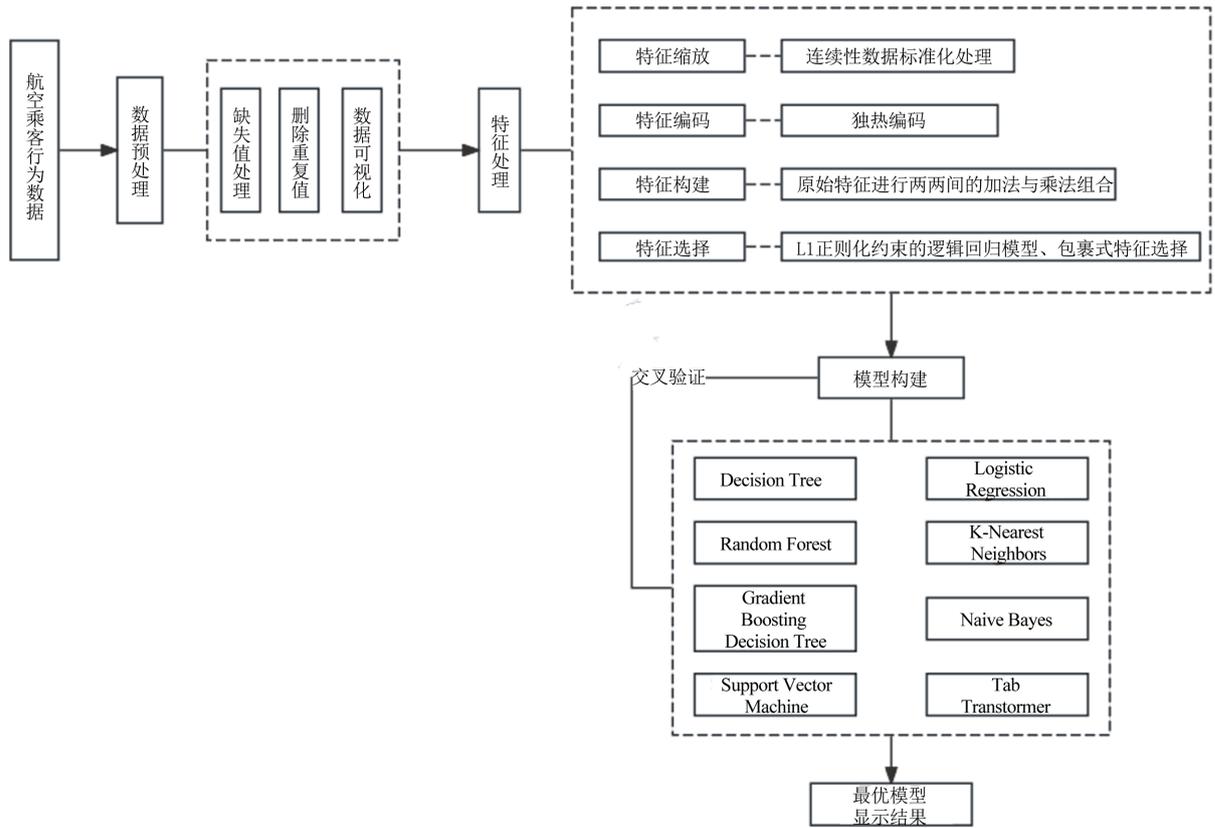


Figure 2. Technology roadmap
图 2. 技术路线图

3. 数据概况及特征工程

3.1. 数据描述

本文所使用的数据样本均来源于 Kaggle 网站, 两组数据都是关于 2015 年在美国完成旅行后的航空乘客的问卷反馈。这两个数据集都处于被引用下载的前列, 并且数据特征名称也相同, 故将两组数据合并研究, 增强可靠性。因为两份数据的数据特征名是相同的, 并且数据的内容也是打分数据, 取值范围在 0~5 之间, 所以可以根据相同名称直接进行合并。数据包括机上 Wi-Fi 服务、出发/到达时间便利性、在线预订便利性、登机口位置等 14 个服务特征, 提供了用户的原始满意度评分: 包括从 0~5 六个等级的

评分, 这些都是航空公司服务的典型指标。同时, 还收集了乘客的特征, 如年龄、性别、航班等级和客户类型。最后, 乘客将他们对飞行的总体满意度分为两类: 满意或不满意/中性。针对各个特征属性的含义进行维度划分, 有服务特征、乘客特征、行程特征、延误特征。其中将服务属性按照时间跨度划分为飞行前、飞行中、飞行后三组。

3.2. 数据清洗

在现实生活中, 采集的原始数据不可避免地存在一些问题, 被称之为“脏数据”, 会对模型的准确性产生一定的影响, 因而需要通过数据清洗来整理数据。在原始数据集中, 包含“id”这样一个表示“序号”的特征, 对于预测模型无实际意义, 予以删除。本文通过箱型图分析数据异常值, 直观地发现存在异常值的特征有飞行距离、出发延迟和到达延迟。通过查阅航空官网的飞行资料, 这种较长的飞行距离是正常存在的, 可以保留, 故未作处理。为防止数据存在交叉重复造成数据冗余, 继续检查重复数据, 发现存在 376 条重复数据, 进行删除。

本文使用 pandas 中的 dropna 方法过滤有缺失值的数据, 缺失率为 0.20%。由于该特征缺失比率很低, 删除这些缺失值所在的行对数据集影响不大, 故予以删除, 最终剩余 233,081 条有效数据, 如表 1 所示。为了进一步的数据分析和机器学习建模, 将分类数据通过编码转换为数值格式。其中对性别、客户类型、旅行类型、舱位等级、满意度进行编码, 转换为 0 或 1。通过编码所有数据类型均为整数型或浮点型, 包括 1 列待预测的目标值满意度和 22 列特征值。“中立或不满意”和“满意”类别分别占有所有数据记录的 56% 和 43%, 比例接近于平衡, 可以认为是平衡数据。后续使用 Python 3.7 中的开源库在 Jupyter Notebook 平台上进行机器学习数据处理、模型构建、评估和相关分析。

Table 1. Basic information situation

表 1. 数据基本情况

序号	特征值名称	Non-Null	数据类型	种类数量
Pre-flight service attributes				
1	Departure/Arrival_time_convenient	233,081	int64	6
2	Ease_of_Online_booking	233,081	int64	6
3	Gate_location	233,081	int64	6
4	Online_boarding	233,081	int64	6
5	On-board_service	233,081	int64	6
6	Checkin_service	233,081	int64	6
During-flight service attributes				
7	Inflight_wifi_service	233,081	int64	6
8	Food_and_drink	233,081	int64	6
9	Seat_comfort	233,081	int64	6
10	Inflight_entertainment	233,081	int64	6
11	Leg_room_service	233,081	int64	6
12	Inflight_service	233,081	int64	6
13	Cleanliness	233,081	int64	6
Post-flight service attributes				

续表

14	Baggage_handling	233,081	int64	5
Passenger Characteristics				
15	Customer_Type	23381	object	2
16	Age	233,081	int64	75
17	Type_of_Travel	233,081	object	2
18	Gender	233,081	object	2
Trip features				
19	Class	233,081	object	3
20	Flight_Distance	233,081	int64	3817
Delayed characteristics				
21	Departure_Delay_in_Minutes	233,081	float64	131
22	Arrival_Delay_in_Minutes	233,081	float64	132
23	satisfaction	233,081	object	2

3.3. 探索性数据分析

为了初步了解与航空满意度有很强相关性的因素，本文计算了斯皮尔曼相关系数，得到特征相关性热图，如图 3 所示，有些因素与整体满意度有很强的相关性，如在线办理登机、旅游类型。

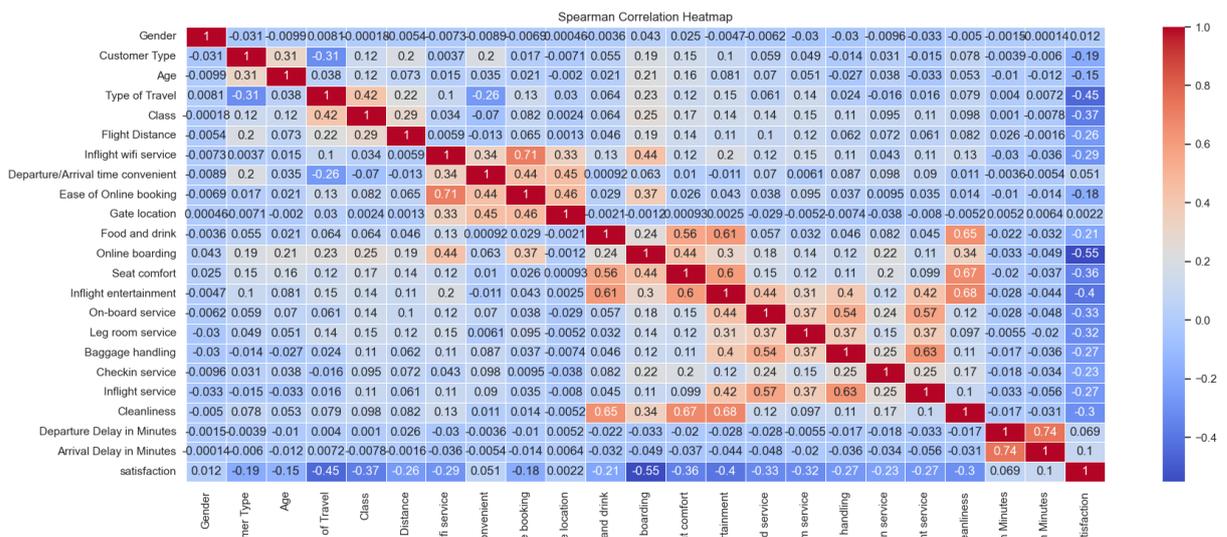


Figure 3. Feature correlation heatmap
图 3. 特征相关性热图

此外，不同因素(如机上 Wi-Fi 服务和在线预订便利性)之间存在相关性，这意味着解释变量之间可能存在共线性。各因素之间的相关性可能归因于航空公司在机上 Wi-Fi 服务和在线预订便利性方面提供良好的服务，而不必要地表示任何因果关系。因此，有必要解决由于特征之间的相关性而产生的潜在偏差。为了解决这一问题，我们采用了基于特征间 Spearman 相关系数的分层聚类方法，对相关度较高的特征进行聚类。结果如图 4 所示，在建模和结果解释中应考虑不同因素之间的相关性，以避免分析偏差。

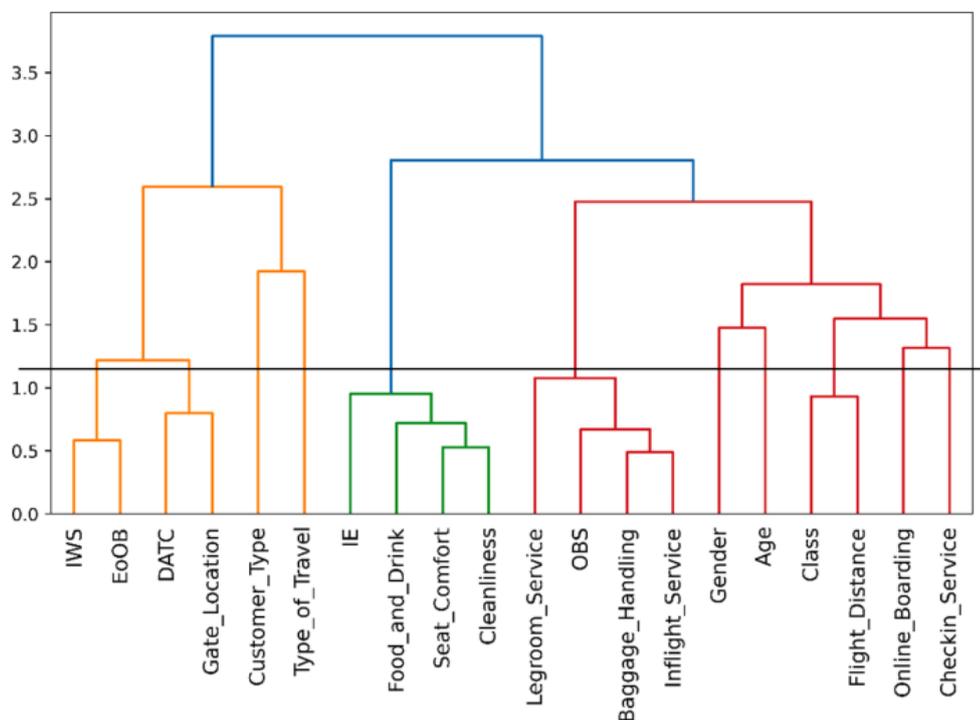
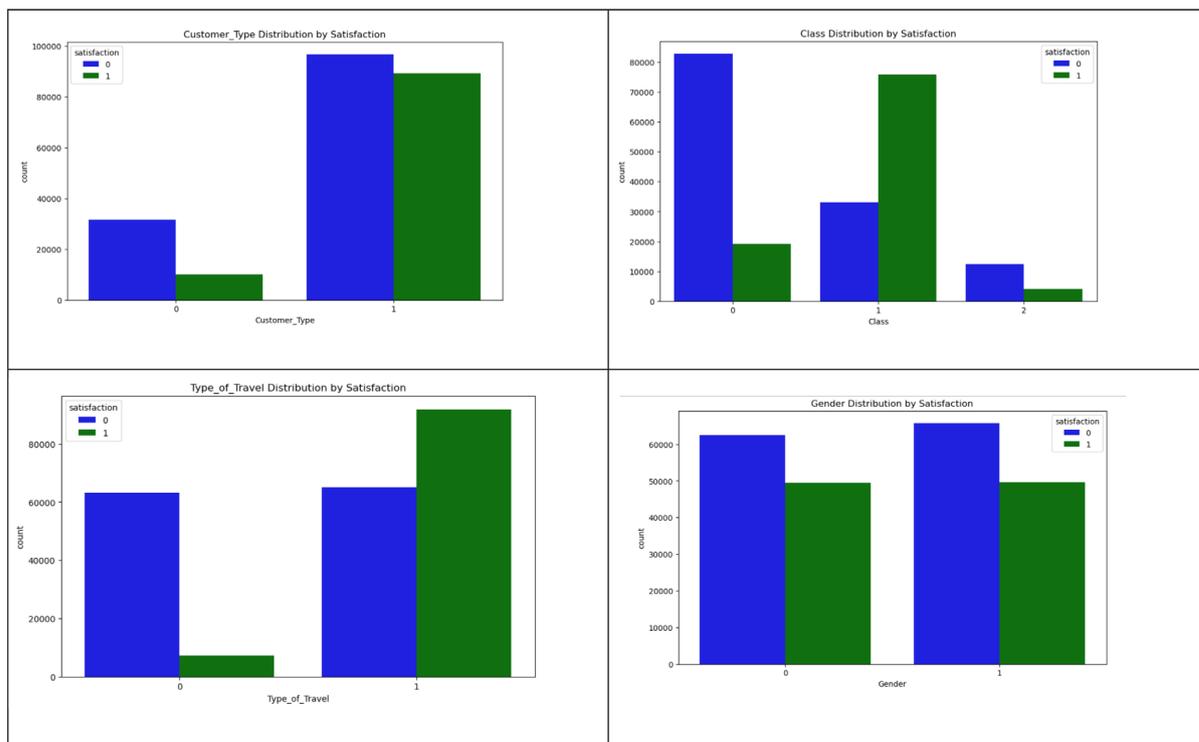
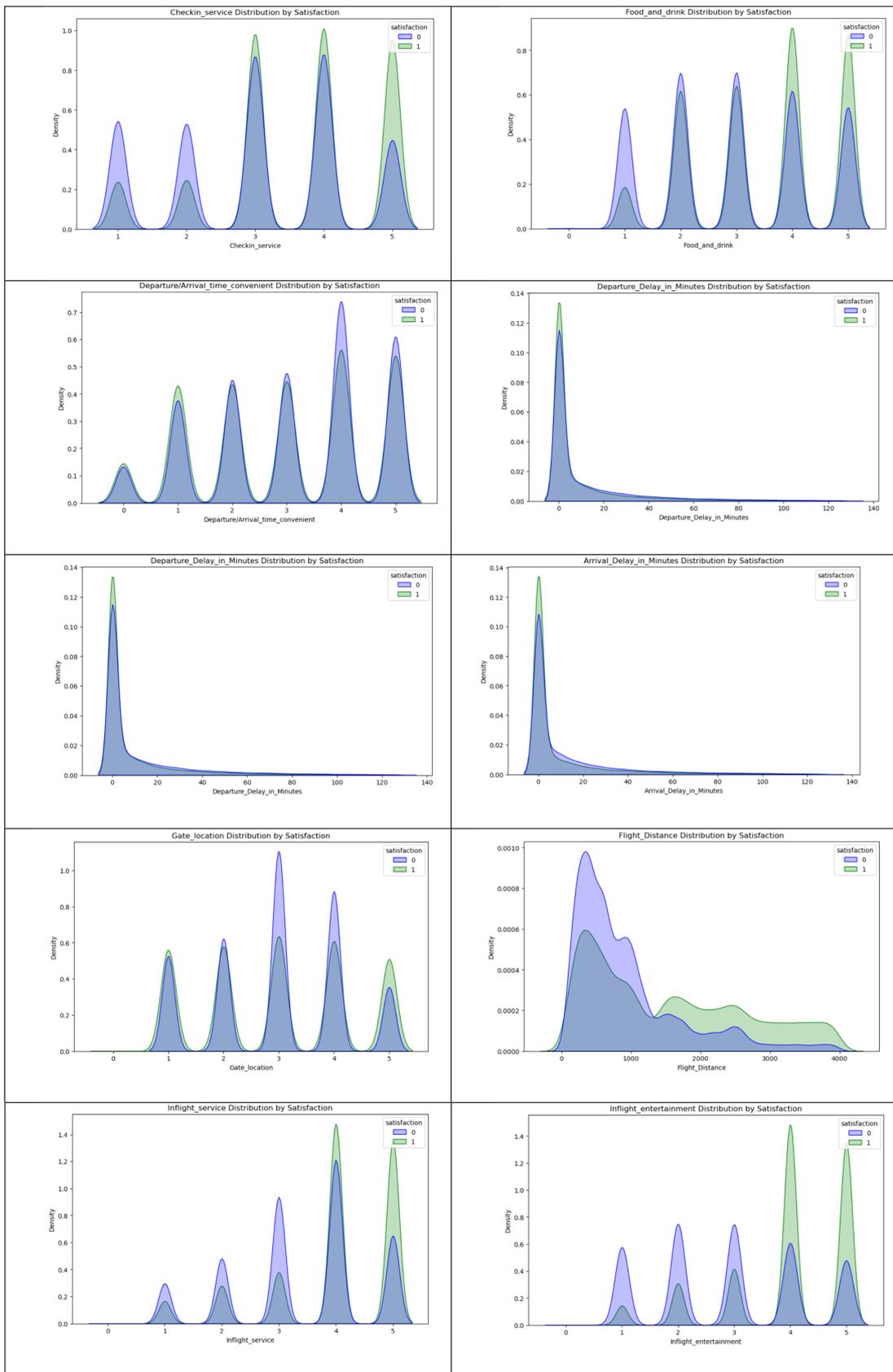


Figure 4. Hierarchical clustering result chart
图 4. 分层聚类结果图

各属性的直方图和核密度估计图如图 5 所示。分类属性在视觉上根据标签进行分解(“不满意或中性”是蓝色的, “满意”是绿色的)。数值属性通过核密度估计可视化, 按标号类进行图分解。





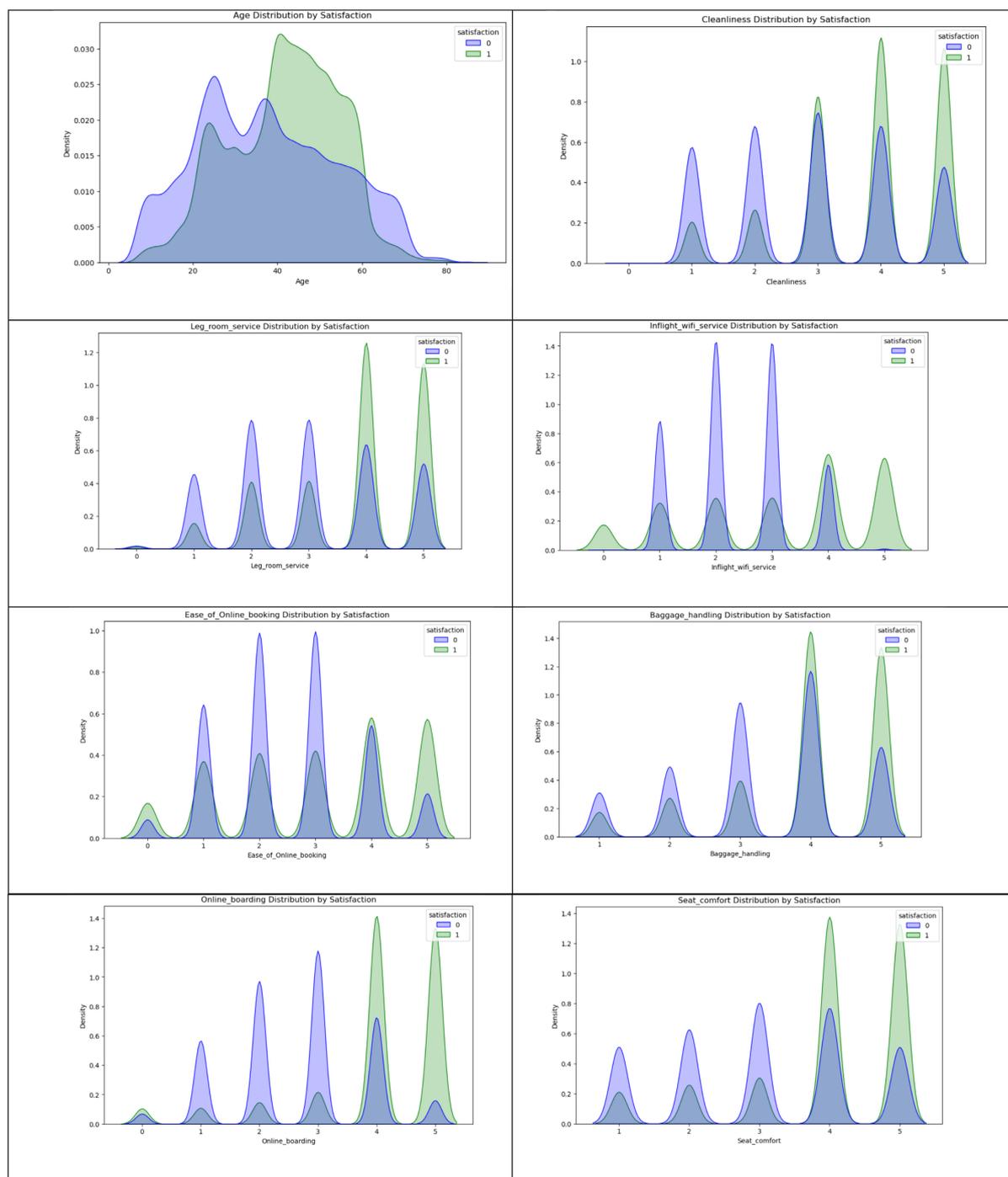


Figure 5. Histogram and kernel density estimation plot for each property
图 5. 每个属性的直方图和核密度估计图

3.4. 数据编码

在本研究中, 对分类属性进行了合适的数值编码, 以便于机器学习模型的训练和分析。

本文的编码处理如表 2 所示, 编码后, 自变量个数由 11 个变为 14 个, 样本数为 7831 个。对于数值型变量 “Age”、“Flight_Distance”、“Inflight_wifi_service”、“Departure/Arrival_time_convenient”,

则不需要进行编码。

Table 2. Feature coding correspondence

表 2. 特征编码情况

	变量	赋值
1	Gender	Male = 0, Female = 1
2	Customer_Type	Disloyal Customer = 0, Loyal Customer = 1
3	Type_of_Travel	Personal Travel = 0, Business travel = 1
4	Class	Eco = 0, Business = 1, Eco Plus = 2
5	Satisfaction	Satisfied = 0, Neutral or Dissatisfied = 1

3.5. 特征构造

特征构造是指利用原始特征来构造新的特征,一般要结合数据实际情况,构造出更有意义的新特征,从而提升模型的预测表现能力或更好地解释模型。为捕捉特征间的非线性关联与交互效应,本研究采用特征交叉方法,对数据中所有 22 个原始特征进行两两间的加法与乘法组合,共生成 462 个新特征。其中,加法项用于建模特征间的线性协同效应,乘法项用于刻画非线性交互关系,以此增强模型对复杂数据模式的表达能力。

利用特征选择方法,可以减少无效特征,降低模型训练的难度和复杂度,有效避免模型过拟合,缩短模型的运行时间,从而提高模型的学习效率。构建后的数据集属于高维数据,因此必须进行特征选择,将与乘客满意度无关的特征、冗余特征删除,从而获得更好的性能。本文采用两种方法,第一种基于 L1 正则化约束的逻辑回归模型,有效剔除与目标变量弱关联的冗余特征,另一种方法采用 Wrapper 的前向选择法,从数据集中筛选出不必要的特征。

3.6. 特征筛选

逻辑回归(Logistic Regression)是一种用于分类任务的机器学习算法。L1 正则化约束的逻辑回归模型是在传统逻辑回归模型的基础上,通过在损失函数中添加 L1 正则化项来实现模型的正则化,防止过拟合,并且能够进行特征选择。通过该方法最终得到 359 个特征。

Table 3. Comparison of two modeling results of the LightGBM algorithm

表 3. LightGBM 算法的两种建模结果对比

	方式 1 的特征集	方式 2 的特征集
Accuracy	96.01%	95.92%
AUC	96.12%	96.04%
Precision	95.84%	95.73%
Recall	96.12%	96.04%
F1-score	95.96%	95.86%

机器学习包裹式特征选择通过评估特征子集对模型性能的影响来筛选特征,序列前向选择法是其中的重要算法。该算法从空特征集合出发,每次往集合中添加一个特征,并重新训练模型、评估性能。每一步都会考量未入选的特征,将加入后模型性能最优的特征纳入集合。直至满足停止条件,如特征数达

阈值或模型性能提升不显著。其优势在于能考虑特征组合效应, 通过重训模型评估子集性能, 相比过滤式方法, 通常能得到更优的特征子集。通过该方法最终得到 206 个特征。

为了从两种方法中筛选出质量较高的特征集, 本文以 LightGBM 模型作为筛选的基准模型, 将上述方法得到的两个特征集带入该模型, 得到分类结果, 用 Accuracy、AUC、Precision、Recall、F1-score 等评价指标对比选择质量较高的特征集, 见表 3。

实验表明方式 1 构建和处理的特征集质量更高, 后续用方式 1 的数据。

4. 建模与评估

本文的研究目的是根据航空公司乘客行为数据来预测乘客对本次的乘坐服务是否满意, 并且找出哪些特征(包括原始特征和构造特征)对乘客满意度影响较大, 从而为乘客制定个性化的服务需求, 提升航空公司的服务质量。本小节基于最优特征集建模然后进行模型对比分析, 即将八种算法所得的结果进行对比分析。

4.1. 划分数据集

本文使用 scikit-learn 库里的 `train test split()` 函数随机地划分数据集, 80% 的数据用来训练, 20% 的数据用来测试。其中, 训练集包含 186,465 条数据, 测试集包含 46,616 条数据。

4.2. 网格搜索法调参

在机器学习中, 超参数是模型训练前需人工设定的参数, 直接影响模型效果。不当选择会导致欠拟合或过拟合, 因此参数调优至关重要。网格搜索是常用的调参方法, 通过遍历给定参数组合寻找最优解, 确保结果接近全局最优, 但计算成本高。本文采用分步优化策略: 先筛选关键参数, 按影响程度逐步调优。初始设定较大范围, 逐步缩小直至确定最优值, 以平衡效率与精度。

下面以 GBDT 算法为例, 展示调参过程, 其他七种算法调参步骤与此过程类似, 下文不再赘述。

GBDT 算法的参数包括 boosting 框架参数和弱分类器参数。boosting 框架参数需要调节的参数主要有: `nestimators`、`learningrate` 和 `subsample`。弱分类器需要调节的参数有: `maxdepth`、`min samples split` 和 `min samples leaf`。

具体调参步骤如下:

第一步, 优化迭代次数, `nestimators` 设置取值范围为 [150, 300), 步长为 10。将 `learning rate` 设置为一个较小的值 0.1, `max features` 设置为 “sqrt”, 通过网格搜索, 迭代次数的最优值为 240。

第二步, 优化 `max_depth` 和 `min_samples_split`, 将其取值范围分别设置为 [3, 16), 步长为 2, 通过遍历这些值, 得到最优值为 `max depth = 10`, `min samples split = 800`。

第三步, 由于内部节点分割所需最小样本数 `min samples split` 和叶子节点最少样本数 `min samples leaf` 之间存在关联, 所以需要再次同时优化这两个参数。经过网格搜索, 得到其最优值为 `min samples split = 1200`, `min samples leaf = 80`。

第四步, 优化 `max features` 参数, 先设置其遍历范围为 [10, 22), 步长为 2, 得到最优值为 10。

第五步, 优化 `subsample` 参数。其默认值是 1, 即不使用子采样, 但其值如果过小, 就会增加样本拟合的偏差, 通过搜索得到最优值为 0.8。

第六步, 可以通过减半学习速率、加倍迭代次数来增加我们模型的泛化能力, 即取 `learning rate = 0.05`, `n_estimators = 580`, 但学习速率不宜过小。

经过上述调参步骤后, GBDT 算法的各评价指标值如表 4 所示。

Table 4. The GBDT algorithm predicts the result after parameter tuning
表 4. 调参后的 GBDT 算法预测结果

评价指标	数值
Accuracy	96.71%
AUC	96.62%
Precision	95.94%
Recall	96.52%
F1-score	96.23%

4.3. 模型构建与对比分析

利用数据预处理后的原始特征集, 即 22 个预测变量、1 个目标变量进行建模, 同样使用上述网格搜索法进行调参, 然后采用 GBDT 进行满意度预测, 得到结果后, 与构建新特征的 GBDT 算法进行对比, 结果如表 5 所示。

Table 5. Comparison of the original feature set and the constructed new feature set results
表 5. 原始特征集和构建后的新特征集结果对比

评价指标	原始特征集	构建的新特征集
Accuracy	94.69%	96.71%
AUC	94.22%	96.62%
Precision	92.62%	95.94%
Recall	93.82%	96.52%
F1-score	94.17%	96.23%

可以看到, 原始特征集进行建模得到的准确率、精确率、AUC 值、Recall 以及 F1 score 都低于构建的新特征集, 说明本文的改进思路是可行的, 通过使用特征构造和特征选择算法得到的筛选特征集比原始特征集有更好的表现能力, 可以有效提高算法的预测精度和泛化能力。

Table 6. Comparison of prediction results of eight models
表 6. 八个模型预测结果对比

	Accuracy	Precision	AUC	Recall	F1-score
DT (决策树)	0.9591	0.9651	0.9572	0.9423	0.9525
RF (随机森林)	0.9654	0.9594	0.9562	0.9452	0.9623
GBDT (梯度提升决策树)	0.9675	0.9544	0.9531	0.9411	0.9392
SVM (支持向量机)	0.9562	0.9455	0.9442	0.9383	0.9318
LR (逻辑回归)	0.9374	0.9364	0.9354	0.9310	0.9303
KNN (K 近邻算法)	0.9172	0.9163	0.9157	0.9076	0.9117
NB (朴素贝叶斯)	0.9475	0.9474	0.9454	0.9403	0.9425
Transformer	0.9719	0.9734	0.9669	0.9589	0.9638

本文在特征对比得到结果后, 再次利用八种机器学习算法进行了结果比较, 表 6 是经过网格搜索调

参后的各模型评价指标的结果, 包括准确率、精确率、召回率、F1-score 及 AUC 值。可以发现, 在这八种机器学习算法对于乘客是否满意的预测结果中, 除了 KNN 和逻辑回归, 其他模型的准确率都在 95% 以上, 其中 Transformer 算法的准确率、精确率、F1-score 以及 AUC 值是所有模型中最高的, 分别为 0.9719、0.9834、0.9669、0.9489、0.9638, 集成算法占据优势, GBDT 和 RF 紧随其后, 体现集成学习的稳定性, 传统模型中 SVM 和 NB 表现中等, LR 和 KNN 相对较弱, 可能与数据特征或参数调优有关。根据表格信息, 本文做了八个机器学习模型预测结果的折线图, 如图 6 所示。

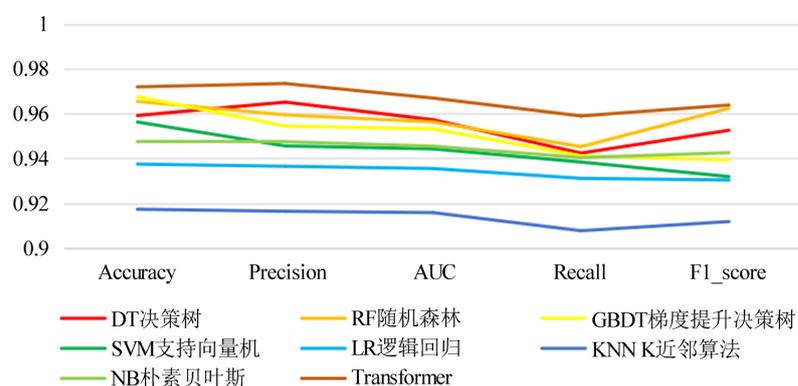


Figure 6. Eight model predictions compared to line charts
图 6. 八个模型预测结果对比折线图

利用随机森林算法对原始特征进行重要性排序, 结果如图 7 所示, 可以看出, 在所有原始特征里, 最重要的特征是线上登机, 其次是机上 Wi-Fi 服务, 另外座椅舒适度、饮食对乘客满意度的影响也比较大, 性别这个特征对乘客满意度几乎没有影响。在线登机和机上 Wi-Fi 服务的便捷性反映了数字化服务的重要性。

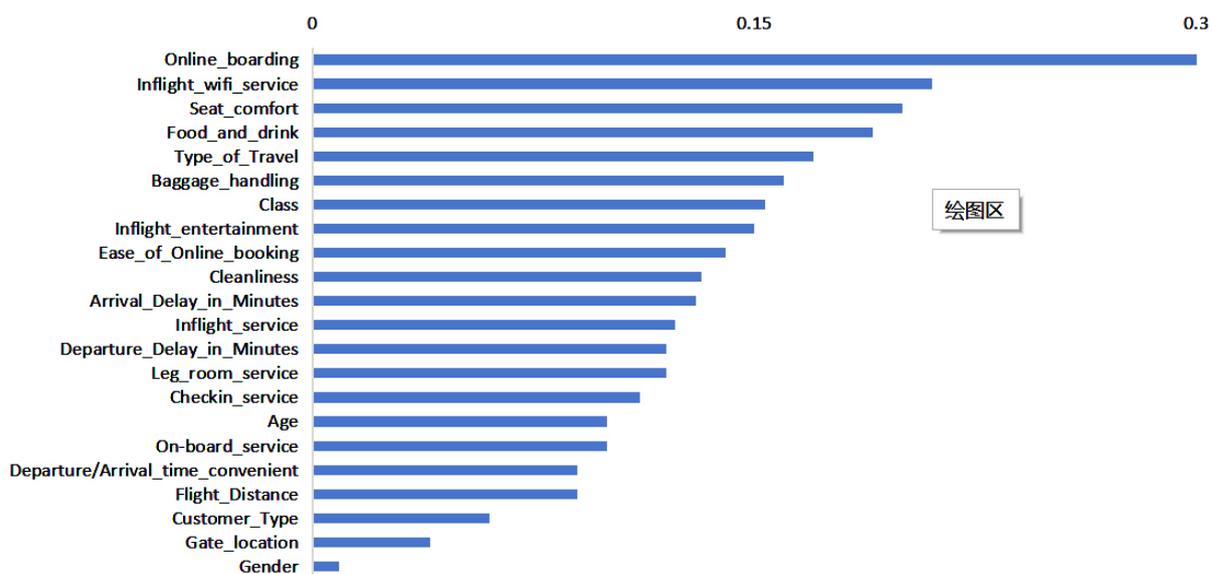


Figure 7. Random forests rank the importance of the original features
图 7. 随机森林对原始特征的重要性排序

5. 结论

根据上文的特征重要性排序结果, 对组合特征进行研究, 发现 Online_boarding 和 Inflight_wifi_service

始终是最重要的特征。并且随着乘客对线上登机 and 无线服务满意度的提高, 他们对航空服务满意程度也会更高。所以航空公司要重点关注那些喜欢线上登机和机上无线服务的乘客。

通过推出差异化的舱位, 可以满足不同乘客的个性化需求。因此, 参照全服务航空的做法, 航空公司可以将经济舱再划分为两种类型, 以供乘客自由选择。第一种乘客强烈需求无线网络, 他们可以选择乘坐有高质量网络服务的机舱, 第二种乘客对无线网络需求没那么强, 他们可以选择乘坐具有丰富线下娱乐活动的机舱或者座椅更舒适的定制舱。这样可以实现对乘客需求的精准划分, 从而提高乘客满意度, 同时还能提高航空公司的运行效率。

线上登机服务已成为影响乘客体验的最关键因素。随着数字化进程加速, 这种便捷的登机方式尤其受到年轻旅客的青睐。航空公司应当以此为契机, 持续优化和扩展线上服务平台, 开发更多智能化功能来提升用户体验。

针对老年等数字弱势群体, 航空公司需要制定差异化服务策略: 一方面简化线上操作界面, 提供清晰的使用指引; 另一方面保留传统服务渠道, 并引入 AI 语音助手、视频指导等智能辅助手段, 帮助他们逐步适应数字化服务。

从长远发展来看, 航空公司要把握技术变革机遇, 深度融合 5G、大数据、人工智能等创新科技, 构建“线上 + 线下”的智慧服务体系。通过技术创新和服务升级双轮驱动, 既满足数字化原生代旅客对效率的追求, 又兼顾特殊群体的服务需求, 最终实现服务品质和乘客满意度的全面提升。

参考文献

- [1] 毛曼, 朱金福. 航空公司顾客满意度测评模型及实证研究[J]. 交通运输系统工程与信息, 2010, 10(5): 201-207.
- [2] Park, E., Jang, Y., Kim, J., Jeong, N.J., Bae, K. and del Pobil, A.P. (2019) Determinants of Customer Satisfaction with Airline Services: An Analysis Off Customer Feedback Big Data. *Journal of Retailing and Consumer Services*, **51**, 186-190. <https://doi.org/10.1016/j.jretconser.2019.06.009>
- [3] Huang, Y. (2023) Low-Cost Airlines Not So Low-Cost—Exploring the Relationships among Travel Motivation, Service Quality and Satisfaction: The Moderating Roles of Perceived Value. *Research in Transportation Business & Management*, **49**, Article ID: 101008. <https://doi.org/10.1016/j.rtbm.2023.101008>
- [4] Shin, R., Park, J. and Choi, D. (2024) Disruptive Passenger Behavior Impact on Overall Service Experience: An Appraisal Theory Perspective. *Sustainability*, **16**, Article 2773. <https://doi.org/10.3390/su16072773>
- [5] Gao, K., Yang, Y. and Qu, X. (2021) Examining Nonlinear and Interaction Effects of Multiple Determinants on Airline Travel Satisfaction. *Transportation Research Part D: Transport and Environment*, **97**, Article ID: 102957. <https://doi.org/10.1016/j.trd.2021.102957>
- [6] Park, S., Kim, M., Kim, Y. and Park, Y. (2022) A Deep Learning Approach to Analyze Airline Customer Propensities: The Case of South Korea. *Applied Sciences*, **12**, Article 1916. <https://doi.org/10.3390/app12041916>
- [7] Pranav, R. and Gururaja, H.S. (2023) Explainable Stacking Machine Learning Ensemble for Predicting Airline Customer Satisfaction. In: Kumar, S., Sharma, H., Balachandran, K., Kim, J.H. and Bansal, J.C., Eds., *Third Congress on Intelligent Systems*, Springer, 41-56. https://doi.org/10.1007/978-981-19-9225-4_4
- [8] Noviantoro, T. and Huang, J. (2022) Investigating Airline Passenger Satisfaction: Data Mining Method. *Research in Transportation Business & Management*, **43**, Article ID: 100726. <https://doi.org/10.1016/j.rtbm.2021.100726>
- [9] Jiang, X., Zhang, Y., Li, Y. and Zhang, B. (2022) Forecast and Analysis of Aircraft Passenger Satisfaction Based on RF-RFE-LR Model. *Scientific Reports*, **12**, Article No. 11174. <https://doi.org/10.1038/s41598-022-14566-3>
- [10] Song, C., Ma, X., Ardizzone, C. and Zhuang, J. (2024) The Adverse Impact of Flight Delays on Passenger Satisfaction: An Innovative Prediction Model Utilizing Wide & Deep Learning. *Journal of Air Transport Management*, **114**, Article ID: 102511. <https://doi.org/10.1016/j.jairtraman.2023.102511>