



XXXX

客户满意度分析报告生成方法研究

徐**¹

摘要: 在通信行业, 保证高水平的客户满意度是保持竞争优势的关键, 客户满意度预测因受样本稀缺或数据不完整情况, 导致预测准确率低, 同时, 模型预测结果的可解释性差, 无法为不满客户进行差异性关怀提供有效的数据支撑。提出一种结合零样本学习的客户满意度预测方法, 通过ETS和t-SNE提取的辅助特征为零样本学习提供了必要的支持, 使得模型能够在没有大量标注数据的情况下, 依然保持较高的预测准确度, 同时, 采用SHAP清晰地展示每个特征对预测结果的贡献, 帮助企业识别并定位客户不满意的关键因素。采集东南某省电调客户满意度数据对模型进行验证, 结果表明, 该方法准确率提升近9%左右, 基于不满客户的解释性分析结果进行客户关怀的成功率达81%以上。

关键词: 客户满意度预测; 可解释性分析; ETS分解; t-SNE降维

中图分类号: TP393

文献标志码: A

doi: 10.11959/j.issn.1000-0801.

Research on the Method of Generating Customer Satisfaction Analysis Reports

Xu Peicai

China Mobile Group Design Institute Co., Ltd. Fujian Branch, Fuzhou 350000, China

Abstract: In the communication industry, ensuring a high level of customer satisfaction is the key to maintaining competitive advantage. Customer satisfaction prediction has low accuracy due to the scarcity of samples or incomplete data. At the same time, the interpretability of model prediction results is poor, which cannot provide effective data support for differentiated care for dissatisfied customers. This article proposes a customer satisfaction prediction method that combines zero sample learning. The auxiliary features extracted by ETS and t-SNE provide necessary support for zero sample learning, enabling the model to maintain high prediction accuracy without a large amount of labeled data. At the same time, SHAP is used to clearly demonstrate the contribution of each feature to the prediction results, helping enterprises identify and locate key factors of customer dissatisfaction. Collecting customer satisfaction data from a certain province in Southeast China to validate the model, the results showed that the accuracy of customer satisfaction prediction in this paper improved by nearly 9%, and the success rate of customer care based on explanatory analysis of dissatisfied customers reached over 81%.

Key words: customer satisfaction prediction, Interpretability analysis, ETS decomposition, T-SNE dimensionality

收稿日期: 2025-10-28; 修回日期: 2025-12-03

通信作者: 徐***,

基金项目: 国家自然科学基金资助项目 (No.xxxxxxx)



reduction

(1. *****公司, 福建 福州 350000)

1 引言

在5G飞速发展的时代背景下, 通信企业发展将又向前迈出重大一步, 我国通信行业竞争日趋激烈。而客户流失是直接影响运营商企业底线的关键指标[1], 客户的流失破坏了客户获取成本与保留费用之间的平衡[2], 影响了电信公司的整体盈利[3]。针对电信行业客户流失预测中存在的挑战, Coussement 等人[4]探索了改进的方法。Liu, Y. 等人[5]采用数据平衡处理和集成学习预测电信行业的客户流失。M. R. Mohaimin 等人[6]采用机器学习算法和高级数据分析来揭示客户不满或流失意图的模式和趋势。众多研究已深入探讨了电信客户流失问题, 并采用多种机器学习和人工智能神经网络技术进行客户流失预测[7]-[11]。在通信行业, 保证高水平的客户满意度是保持竞争优势的关键, 通过预测客户满意度, 明确客户满意度的重要性, 了解目前移动运营商在产品和服务方面存在的一些问题, 针对具体的问题提出具体的解决方案, 做好老客维护稳定, 提高新老客户满意度, 从而取得更好地发展, 进一步巩固好市场的领先优势。

客户满意度预测是一种通过分析客户多维特征数据来评估其对产品或服务满意程度的技术, 已经广泛应用于提升客户体验和优化企业服务策略。2019年, Yingying Wu[12]等人则将客户满意度预测应用到客户搜索领域。吕品等[13]研究了在线产品评论客户满意度综合评价。马艳[14]研究了电力行业客户满意度测评方法。但对电信客户满意度预测研究的文献相对较少, 且缺少对满意度预测结果可解释性的描述。模型可解释性是以可理解的术语澄清或传达模型预测的能力[15-16]。LLM在意图理解和文本生成等方面表现显

著[17-18]。

为解决电信客户满意度学习样本稀缺、可解释性差的问题。本文提出结合零样本学习的客户满意度预测方法, 通过时间序列分解(ETS)从客户的时间序列数据中提取出长期趋势和季节性模式, 这些趋势性特征帮助模型捕捉到客户满意度变化的动态规律。然后, 通过降维技术(t-SNE)将高维数据映射到低维空间, 使得数据的局部结构更加清晰, 进一步增强了模型的特征表示能力, 基于ETS和t-SNE提取的辅助特征为零样本学习提供了必要的支持, 使得模型能够在没有大量标注数据的情况下, 依然保持较高的预测准确度。同时, 本文采用SHAP(SHapley Additive exPlanations)技术清晰地展示每个特征对预测结果的贡献, 帮助企业识别并定位客户不满意的關鍵因素。将这个技术集成并作为Agent工具供于大模型调用此能力, 从而使得业务人员能更好的使用这个工具, 从而了解客户的不满原因后进行精准优化与关怀。

2 相关工作

2.1 分类预测分析

电信客户满意度预测属于分类任务, 如朴素贝叶斯(NB)[19]、决策树(DT)[20]、支持向量机[21]、随机森林(RF)[22], 自适应增强(Ada Boost)[23], XGBoost[24]和LightGBM[25]等机器学习算法被用于分类预测, 神经网络(ANN)也常被应用与分类任务, 神经网络由三个主要处理层组成: 输入层、隐藏层和输出层。许多ANN算法被用于分类任务, 但多层感知器(MLP)[26]、卷积神经网络[27]和长短期记忆神经网络(LSTM)[28]在文献中最常用于分类预测。

2.2 模型可解释

为了强调模型透明度的概念，研究人员利用可解释人工智能(XAI)打破学习的“黑匣子”，使其变成一个透明的“玻璃盒”[29]，对模型输出的基本原理进行解释。XAI通过可解释模型构建决策路径追溯与人性化解释生成技术，实现算法决策过程的透明化与人类认知对齐。其核心突破在于高保真特征归因、因果推理框架和自然语言解释生成，使复杂模型在保持预测精度的同时具备决策逻辑的可视化表达能力。当XAI方法阐明单一预测时（如Shapley值、LIME），它被归类为局部方法；当它解释完整模型时（如特征重要性技术、SHAP），它则被归类为全局方法[30]。SHAP有助于对比解释，拥有强大的理论支持，并提供在特征值之间公平分配的全面解释。缺点：SHAP需要大量的计算资源。LIME使用局部替代模型来解释个体预测（局部解释）。

2.3 LLM外部工具调用

最近，大语言模型（LLMs）因其在技术上的突破，衍生出了一系列前所未有的强大能力，使其从纯粹的文本处理工具演变为能够理解、推理并与世界交互的通用智能体。LLMs通过链式推理（Chain-of-Thought, CoT）和其进阶技术（如思维树 Tree of Thought），能够将复杂问题分解为多个中间步骤，逐步推导出最终答案[31]。凭借Transformer架构的核心——自注意力机制，LLMs能够精准捕捉上下文中的细微语义和长距离依赖关系。它们不再简单地进行关键词匹配，而是真正理解语言的意图、情感和隐含信息。它们能够生成流畅、连贯且富有创造性的文本，风格和内容均可通过提示（Prompt）精准控制[32]。LLMs也具备了与外部工具和环境交互的行动能力。通过函数调用（Function Calling）或API调用，LLMs可以成为连接数字世界的“智能中枢”，将自然语言指令转化为具体操作[33]。

3 满意度分析报告生成方法

本文通过收集大量客户行为数据及相关的满意度标签，采用时间序列分解（ETS）技术对客户的历史数据进行趋势分析和季节性分解，从中提取出能够反映客户满意度动态变化的关键信息。这些从ETS提取的趋势性特征和季节性波动，作为辅助信息，提供了对客户行为变化的深刻理解。此外，本文结合了降维技术（t-SNE）将高维数据降维到低维空间，提取出更加简洁且有区分度的特征。t-SNE提供的数据表示不仅有助于优化特征结构，还能为模型在面对新客户时提供更加准确的推理能力。通过将ETS提取的趋势性信息与t-SNE降维得到的特征结合，这些辅助信息为模型提供了重要的上下文支持，使得模型能够在缺少大量标注数据的情况下，依靠已有的知识进行有效推理，从而提高了在小样本或未标记数据环境中的预测准确性。为了进一步提高模型的可解释性，本文采用SHAP技术，利用其提供的特征重要性评分对预测结果进行详细解析，帮助企业精准识别导致客户不满意的关键因素。

在本文的实现过程中，采用Agent智能体作为核心组件，负责执行所有的预测、分析和解释任务。该Agent能够嵌入到大语言模型中，作为智能决策和咨询模块，提供实时的客户满意度预测和解释功能。

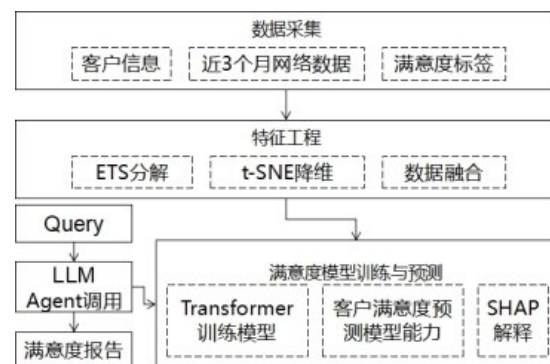


图1 客户满意度分析报告生成流程图



3.1 数据收集及数据分割

本文满意度标签数据收集主要通过两种方式采集：一是利用人工智能（AI）外呼技术，二是通过人工客服电话进行访问。特征维度收集包括客户的年龄、网龄、ARPU值（每客户平均收入）等基本信息。此外，还包括4G和5G网络环境下的上网质量数据，例如4G上网质差占比和5G上网质差占比，这些数据能够反映客户在不同网络环境下的实际体验，并为满意度预测提供时序信息。为了全面了解客户行为，本文通过收集每个客户过去12个月的行为数据，捕捉其网络使用的长期变化趋势。

数据收集完成后，本文通过计算每个客户数据的缺失率，识别并剔除缺失数据较多的记录。具体而言，若某客户数据的缺失率超过30%，则删除该客户的全部数据记录。此外对于连续数值的异常离散点进行剔除。通过这一数据收集和清洗流程，不仅确保了数据能够真实反映客户行为和满意度变化，还保证了数据的准确性和可靠性，为后续的分析建模提供了坚实的基础。

3.2 客户满意度预测模型训练

在客户满意度预测场景中，针对无标签新客户（Unseen Classes / Instances），传统监督学习面临失效风险。本文提出的零样本学习（ZSL）框架，核心在于构建一个共享的语义嵌入空间（Semantic Embedding Space），将原始特征空间与满意度标签空间进行对齐。

定义原始特征空间为 $\mathcal{X} \subseteq \mathbb{R}^d$ ，标签空间为 $\mathcal{Y} = \{0,1\}$ （不满意 / 满意）。在训练阶段，对于可见样本 $S_{\text{seen}} = \{(x_i, y_i, a_i)\}_{i=1}^{N_i}$ ，其中 a_i 为通过 ETS 和 t-SNE 提取的辅助特征向量（Auxiliary Attribute Vector）。假设辅助特征 a_i 能够捕捉客户行为的时序动态性与流形结构，充当 ZSL 中的

“语义描述”等。

模型的推断过程构建如下映射函数 $f: \mathcal{X} \times \mathcal{A} \rightarrow \mathcal{Y}$ 。具体而言，我们设计了一个特征融合网络（Feature Fusion Network），将原始特征 x_i 与辅助语义特征 a_i 进行非线性映射：

$$Z_i = \phi(W_x x_i \oplus W_a a_i + b) \quad (1)$$

其中， \oplus 表示特征拼接操作， W_x 和 W_a 为可学习的投影矩阵，旨在将静态特征与时序 / 结构辅助特征映射到同一高维流形 Z 中。在 Z 空间内，模型学习一个兼容性函数（Compatibility Function） $F(z, y)$ ，即使在缺乏直接标签 y_i 的情况下（即零样本场景），模型也能通过已习得的辅助特征 a_{unseen} 的分布规律，计算新样本属于某一满意度类别的概率：

$$\hat{y} = \arg \max_{y \in \mathcal{Y}} P(y | \phi(x_{\text{unseen}}, a_{\text{unseen}})) \quad (2)$$

通过这种机制，ETS 提供的长期趋势信息和 t-SNE 提供的邻域结构信息，弥补了直接监督信号的缺失，实现了从已知客户到未知客户的知识迁移。模型训练过程如下：

(1) ETS 分解

本文通过对收集到的多维客户特征数据进行时间序列分解，将来自 ETS (Exponential Smoothing State Space Model) 的趋势与季节要素作为辅助特征嵌入原始特征空间中。设有 M 个客户，每个客户 i 在时间长度为 n 的观测窗口内具有一组多维特征序列：

其中，每个特征分量 $x_{t,j}^{(i)}$ 均随时间变化形成

$$\{X_t^{(i)}\}_{t=1}^n, X_t^{(i)} = (x_{t,1}^{(i)}, x_{t,2}^{(i)}, \dots, x_{t,d}^{(i)}) \in \mathbb{R}^d \quad (3)$$

一条特征时间序列。对特征 j 的序列 $\{x_{t,j}^{(i)}\}$ 使用 ETS 分解可得:

$$x_{t,j}^{(i)} = T_{t,j}^{(i)} + S_{t,j}^{(i)} + E_{t,j}^{(i)} \quad (4)$$

其中, $T_{t,j}^{(i)}$ 为该维特征的长期趋势, $S_{t,j}^{(i)}$ 表示周期性变化部分, $E_{t,j}^{(i)}$ 为随机残余项。通过对各特征维度的时序数据进行 ETS 分解, 本提案使用 $\{T_{t,j}^{(i)}\}$ 趋势作为零样本学习中辅助信息的一部分。

最后, 本文通过将辅助特征与客户原始特征相结合, 以构造增强后的特征表示: $\tilde{X}^{(i)} = g(X^{(i)}, T^{(i)})$ 。这里 $X^{(i)}$ 为客户原始特征的代表, 该整合后的特征不仅包含客户原有特征分布信息, 还内嵌了其特征在时间尺度上的动态趋势与季节变化, 为模型在缺乏直接标注数据的情形下, 提供可借鉴的时间结构信息, 实现更为稳定和泛化的推断。

(2) t-SNE 降维

本文在获得包含趋势辅助特征的原始特征表示基础上, 引入 t-SNE (t-Distributed Stochastic Neighbor Embedding) 降维技术, 以进一步提取低维、区分度高的特征表示作为辅助信息, 设 $\tilde{X}^{(i)}$ 为客户 i 的增强特征表示, 该特征已包括原始数据与从 ETS 趋势信息中提取的辅助特征。将所有客户的增强特征集合构成高维矩阵:

$$X = [\tilde{X}^{(1)}, \tilde{X}^{(2)}, \dots, \tilde{X}^{(M)}]^T \in \mathbb{R}^{M \times D} \quad (5)$$

其中 M 为客户数, D 为增强后特征的维度。t-SNE 通过构建高维特征分布 (采用条件概率建模近邻关系) 和相应的低维嵌入分布, 然后反复优化低维嵌入坐标, 以使两种分布在统计意义上保持相似。记低维表示为:

$$y = [y^{(1)}, y^{(2)}, \dots, y^{(M)}]^T \in \mathbb{R}^{M \times d'} \quad (6)$$

其中 $d' \ll D$ 为低维空间的维数。通过 t-SNE 获得的低维特征 $y^{(i)}$ 不仅紧密保留了客户之间在高维空间中的近邻结构关系, 还在一定程度上弱化了噪声和冗余特征的影响, 从而使辅助信息更具判别性和可泛化能力。将这些低维特征与客户的其他辅助信息 (如 ETS 趋势特征) 相结合, 模型能够在零样本学习场景下更清晰的识别新客户在特征结构上的位置, 并基于该位置做出更有依据的满意度推断。

(3) Transformer 预测

在模型训练阶段, 本文将包含原始特征、ETS 趋势辅助信息和 t-SNE 低推特征的增强表示 $\tilde{X}^{(i)}$ 作为输入, 经过适当的序列化处理后输入至 Transformer 模型进行分类预测。设 H 为通过 Transformer 编码对输入序列特征进行编码后得到的隐含表示, 最终的预测结果可记为:

$$\hat{y}^{(i)} = f_{cls}(\text{Transformer}(\tilde{X}^{(i)})) \quad (7)$$

其中 $\text{Transformer}(\cdot)$ 表示 Transformer 编码步骤, f_{cls} 为分类头 (如全连接层加软最大化), $\hat{y}^{(i)}$ 为客户 i 的满意度预测结果。

3.3 模型解释模块

本文的模型解释模块利用 SHAP 技术, 将客户满意度预测结果中各特征的贡献度进行显化和解析。设 $\tilde{X}^{(i)}$ 为客户 i 的最终特征表示 (已融合原始特征、ETS 趋势信息及 t-SNE 低维表示), 模型预测结果为 $\hat{y}^{(i)}$ 。SHAP 值通过比较特征加入与移除对预测结果的影响来评估特征的重要性, 可形式化为:

其中 ϕ_j 为第 j 个特征对预测结果的平均边际



$$\phi_j = \sum_{S \subseteq \{1, \dots, d\} \setminus \{j\}} \frac{|S|!(d - |S| - 1)!}{d!} [f(S \cup \{j\}) - f(S)] \quad (8)$$

贡献度， $f(\cdot)$ 为模型预测函数， d 为特征总维度， S 为特征子集。当将此解释机制应用于本提案的客户满意度预测模型时，可直观展示出原始特征对每个客户的影响状况，提供更为透明和可解释的客户满意度分析依据。针对 SHAP (Kernel Explainer) 在高维数据下计算复杂度高的问题，本文在工程实现中采用了 Background Summary 策略。即并非使用全量训练集作为背景数据，而是通过 K-means 聚类从训练集中提取个代表性样本作为背景分布。这在保持解释偏差低于 5% 的前提下，将单次推理的解释耗时降低了约 90%，满足了 Agent 实时响应的业务需求。

3.4 Agent大语言模型调用

本文进一步将上述客户满意度预测与解释模型通过封装为 Agent 的形式嵌入到大语言模型 (LLM) 中，使其作为智能决策和咨询模块对外提供接口。当企业或客户通过大模型接口进行查询时，Agent 能够实时对输入的客户群体数据进行满意度预测、原因分析与解释。

在实现上，本文将数据预处理、ETS 特征提取、t-SNE 降维、Transformer 分类与 SHAP 解释等关键步骤统一集成到 Agent 中。当大模型向 Agent 请求分析特定客户群体的满意度水平时，Agent workflow 如下：

- (1) 接收包含客户特征序列数据的请求；
- (2) 对输入的多维特征序列进行 ETS 分解与 t-SNE 降维，构建增强的特征表示；
- (3) 利用 Transformer 模型对增强特征进行满意度预测，获取各群体对应的满意度评分或分类结果；
- (4) 借助 SHAP 机制对预测结果进行解释，计算各特征对最终决策的贡献度，筛选出影响客

户满意度的关键因素；

(5) 将不满原因及解释性特征信息返回给上层大模型，以自然语言形式给出分析报告和建议。

通过该 Agent 的自动化处理和分析，本提案能让大模型在与企业运营人员或终端客户对话时，直接给出目标客户群体不满原因的清晰描述与洞察，从而为决策者提供数据驱动的参考信息，在零样本场景下快速识别影响满意度的潜在问题，并提出相应的改进方向。

4 数据实验与分析

本文通过采集客户满意度标签和特征数据，通过将 ETS 分解的时间序列趋势特征与 t-SNE 降维得到的结构化信息相融合，采用 transform 算法构建客户满意度预测模型，并借助 SHAP 的可解释性分析手段，采用 Agent 智能体作为核心组件，负责执行所有的预测、分析和解释任务。该 Agent 能够嵌入到大语言模型中，作为智能决策和咨询模块，提供实时的客户满意度分析报告生成能力。

4.1 特征工程

本文数据来源于东南某省移动客户数据，客户信息指标包括：用户星级、年龄、性别、预付费客户等 37 个指标；网络质量所涵盖的特征指标是用户使用中与网络相关的各项指标，如日均附着时延、日均 TAU 时延、日均 DNS 查询时延、4G 上网质差占比、ARPU 值、语音饱和度等 69 个指标。

采用 ETS 分解技术对近 3 个月的网络指标进行分解，如下图所示：

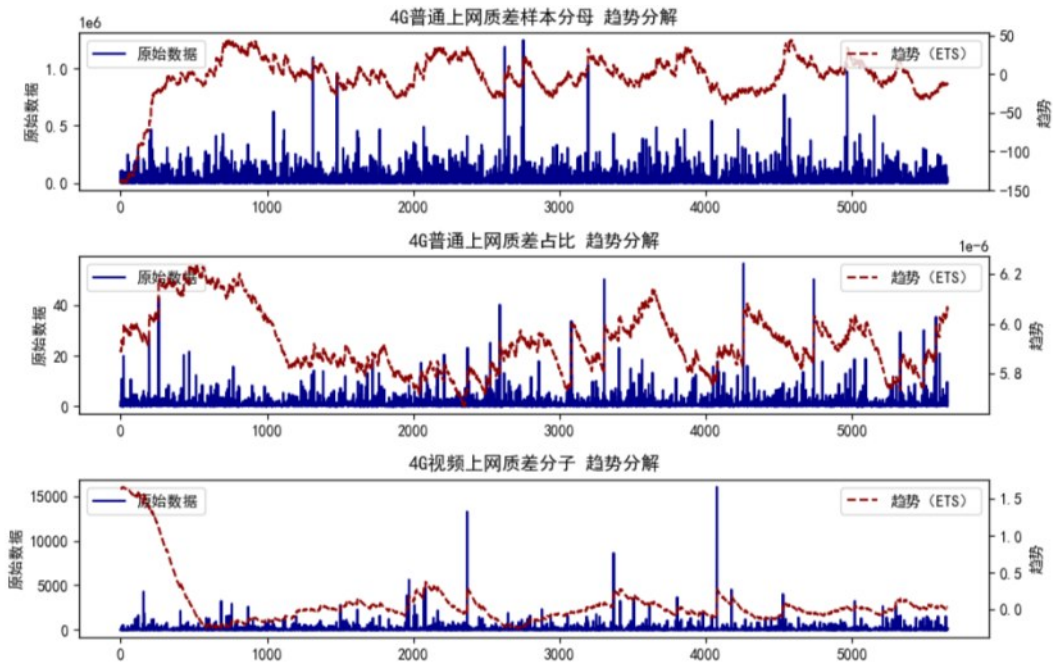


图2 网络指标趋势分解图

4.2 客户满意度预测结果分析

机器学习算法，包括 SVM、XGBoost、LightGBM 和 ANN 等，常用于分类模型训练。通过数据平衡技术（如 SMOT 等）、数据缩放方法（如最小最大缩放）和超参数优化策略（如网格搜索）对各种模型进行优化，提高可泛化性，并纠正客户满意度数据集中存在的类不平衡。使用常规的性能指标（即准确性(A)、精确度(P)、召回率(R)和 F1 评分）对模型进行了评估，以确保对预测效果进行公平的评估。下表3显示各算法模型性能评估指标对比情况。为了全面验证本文提出方法的有效性，除传统机器学习模型外，本实验还引入了主流的预训练语言模型 BERT 和 RoBERTa 进行对比。本文将客户的多维特征序列化为文本 Token 输入，微调预训练模型进行分类任务。表1显示各算法模型性能评估指标对比情况：

由上表可知，通过 ETS 技术对网络指标进行时序分解，并采用 t-SNE 进行数据降维，其模型的性能指标均有显著的提升，其中 Transformer +

表1 不同算法模型性能指标对比情况

模型算法	A	P	R	F1
SVM + SMOTE	0.65	0.66	0.64	0.65
SVM + SMOTE+ ETS + t-SNE	0.69	0.70	0.68	0.69
XGBoost + SMOTE	0.72	0.73	0.71	0.72
XGBoost + SMOTE +ETS+t-SNE	0.76	0.77	0.75	0.76
LightGBM + SMOTE	0.73	0.74	0.72	0.73
LightGBM + SMOTE +ETS+t-SNE	0.77	0.78	0.76	0.77
ANN + SMOTE	0.55	0.59	0.64	0.61
ANN + SMOTE +ETS+t-SNE	0.67	0.65	0.59	0.62
BERT-Base + SMOTE	0.72	0.71	0.67	0.69
BERT-Base + SMOTE +ETS+t-SNE	0.79	0.76	0.78	0.77
RoBERTa + SMOTE	0.71	0.69	0.66	0.67
RoBERTa + SMOTE +ETS+t-SNE	0.76	0.74	0.76	0.75
Transformer + SMOTE	0.77	0.76	0.72	0.74
Transformer + SMOTE +ETS+t-SNE	0.86	0.82	0.84	0.83

SMOTE +ETS+t-SNE 模型的性能最好，用户满意度预测 F1 得分为 0.83。引入 BERT 和 RoBERTa 后，模型性能相比传统 SVM 有显著提升，说明大规模预训练模型的特征提取能力较强。然而，本文提出的 Transformer+ETS+t-SNE 方法在 F1 值上依然优于 BERT 和 RoBERTa 模型。这主要得益于本文方法显式地利用了 ETS 提取的时序趋势



特征和 t-SNE 的流形结构特征，相比于直接将表格数据 Token 化输入 LLM，更能精准捕捉通信客户满意度的特定领域规律。

为了进一步进行可解释性分析，下图展示了各特征维度对用户满意度预测的重要性，很明显可以看出年龄越大，给的评分越低，网龄越大评分也越低。

本文通过该 Agent 的自动化处理和分析，让大模型在与企业运营人员或终端用户对话时，直接给出目标用户群体不满原因的清晰描述与洞察，从而为决策者提供数据驱动的参考信息。如

下展示了该 Agent 的后台思考链。

4 结束语

与现有技术相比，本文在面临数据稀缺和用户多样化场景时能够更加稳健地完成用户满意度预测，将零样本学习策略与 ETS 分解后的时间序列动态特征和 t-SNE 降维所得的结构化特征有机融合，从而在多维度数据下获得更丰富的潜在信息表达。此外，本申请还利用 SHAP 进行解释性分析，为企业决策者提供更加透明且可操作的原因定位能力，使得在数据不足条件下依然能够实

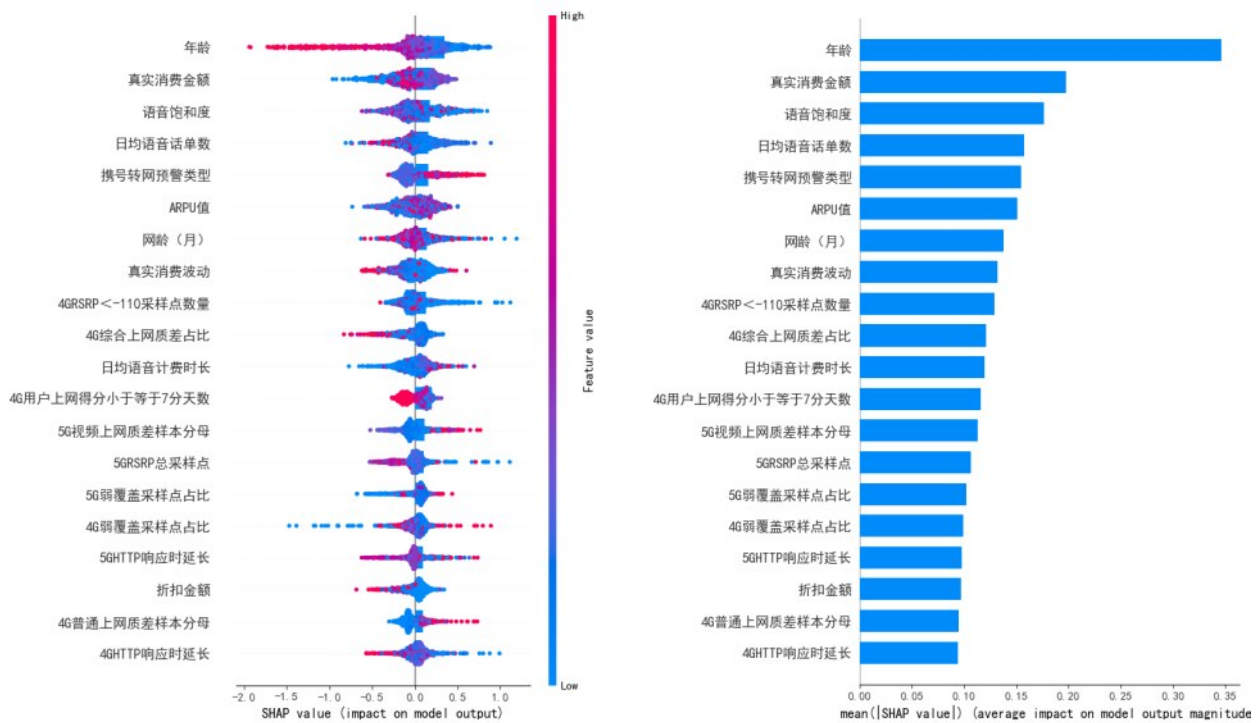


图3 SHAP总体用户重要度示意图

{‘input’: ‘请分析 XXX 用户满意度报告’}

‘output’: ‘用户的不满原因为年龄 35、4GHTTP 响应时延长 11.73、4G 用户视频得分小于等于 7 分天数 10.24、4G 综合上网质差占比 9.08、5G 用户语音得分小于等于 7 分天数 4.92、其它上网不满意次数 4.82、是否流量超套 3.39、5GHTTP 响应时延长 2.94、语音综合感知质差样本占比 2.71、TAU 失败次数 2.68、4G 用户上网得分小于等于 7 分天数 2.35’、4G 用户语音得分小于等于 7 分天数 1.7、5G 综合上网质差占比 0.95、4GTCP 建链失败 0.92、4G 上网感知波动次数 0.85、NR 视频不满意次数 0.83、5G 弱覆盖采样点占比 0.61、视频不满意次数 0.39}

现较高的预测精度和可解释性，同时为后续策略优化提供清晰的依据。

尽管本文实验基于某省的电信数据，结果验证了方法在特定区域的有效性。考虑到不同地域（如内陆与沿海）的用户行为习惯与网络覆盖差异，模型在跨域迁移时可能面临分布偏移（Distribution Shift）挑战。未来的工作将引入迁移学习（Transfer Learning）技术，利用本文提出的 ETS+t-SNE 辅助特征作为域不变特征（Domain-Invariant Features），提升模型在多地域数据上的泛化能力

参考文献：

- [1] Verbeke W., Dejaeger K., Martens D., et al. New Insights into Churn Prediction in the Telecommunication Sector: A Profit Driven Data Mining Approach [J]. *European Journal of Operational Research*, 2012, 218(1): 211-229.
- [2] Ascarza E., Neslin S. A., Netzer O., et al. In pursuit of enhanced customer retention management: Review, key issues, and future directions [J]. *Customer Needs and Solutions*, 2018, 5: 65-81.
- [3] Xie Y., Wang Y., & Lin Y. Telecom Customer Churn Prediction in Big Data: A Cognition Computing Approach [J]. *IEEE Access*, 2016, 4: 3773-3782.
- [4] Coussement K., Lessmann S., & Verstraeten G. A comparative analysis of data preparation algorithms for customer churn prediction: A case study in the telecommunication industry [J]. *Decision Support Systems*, 2017, 95: 27-36.
- [5] Liu Y., Sun L., & Wu J. Customer churn prediction in telecommunications industry using data balance processing and ensemble learning [J]. *Computers, Materials & Continua*, 2018, 56(1): 1-19.
- [6] M. R. Mohaimin, B. C. Das, R. Akter et al. Predictive Analytics for Telecom Customer Churn: Enhancing Retention Strategies in the US Market [J]. *Journal of Computer Science and Technology Studies*, 2025, 7(1): 30-45.
- [7] Somya, S. Kaushik, S. Saini, S. et al. Telecom Churn Prediction Using Data Science [J]. *Educational Administration: Theory and Practice*, 2024, 30(5): 11026-11034.
- [8] S. Saha, C. Saha, M. M. Haque, et al. ChurnNet: Deep Learning Enhanced Customer Churn Prediction in Telecommunication Industry [J]. *IEEE Access*, 2024, 12: 4471-4484.
- [9] B. Prabadevi, R. Shalini, and B. R. Kavitha. Customer churning analysis using machine learning algorithms [J]. *International Journal of Intelligent Networks*, 2023, 4: 145-154.
- [10] S. Höppner, E. Stripling, B. Baesens, et al. Profit driven decision trees for churn prediction [J]. *European Journal of Operational Research*, 2020, 284(3): 920 - 933.
- [11] G. Sam, P. Asuquo, and B. Stephen, Customer Churn Prediction using Machine Learning Models [J]. *Journal of Engineering Research and Reports*, 2024, 26(2): 181-193.
- [12] Yingying Wu, Yiqun Liu, Yen-His Richard Tsai, et al. Investigating the Role of Eye Movements and Physiological Signals in Search Satisfaction Prediction using Geometric Analysis [J]. *Journal of Neurosurgery Spine*, 2019, 8: 981-999.
- [13] 吕品, 钟珞, 唐琨皓. 在线产品评论客户满意度综合评价研究 [J]. *电子学报*, 2014, 42(4): 740-746.
- [14] 马艳. 模糊理论与信息熵在电力设计行业顾客满意度测评中的应用 [J]. *计算机系统应用*, 2005, 14(5): 39-42.
- [15] Y. Rong, T. Leemann, Thai-Trang N. et al. Towards Human-Centered Explainable AI: A Survey of User Studies for Model Explanations [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2024, 46(4): 2104-2122.
- [16] C. Özkurt, Transparency in Decision-Making: The Role of Explainable AI (XAI) in Customer Churn Analysis [J]. *Information Technology in Economics and Business*, 2025: 1-15.
- [17] C. Qin, A. Zhang, Z. Zhang, et al. Is ChatGPT a General-Purpose Natural Language Processing Task Solver? [J/OL]. *ArXiv*, .2023[2025-03-01]. <http://arxiv.org/abs/2302.06476>.
- [18] H. Naveed, A. U. Khan, Shi Qiuet al., A Comprehensive Overview of Large Language Models [J/OL]. *ArXiv*, 2023[2025-03-01]. <http://arxiv.org/abs/2307.06435>.
- [19] 任世超, 黄子良. 基于二维信息增益加权的朴素贝叶斯分类算法[J]. *计算机系统应用*, 2019, 28(6): 135-140.
- [20] 韩存鸽, 叶球孙. 决策树分类算法中C4.5算法的研究与改进 [J]. *计算机系统应用*, 2019, 28(6): 198-202.
- [21] 江勋林. 多目标支持向量机及其在少样本故障诊断中的应用 [J]. *计算机系统应用*, 2022, 31(9): 287-293.
- [22] 谢国荣, 郑宏, 林伟圻, 等. 基于改进随机森林算法的停电敏感客户分类[J]. *计算机系统应用*, 2019, 28(3): 104-110.
- [23] A. Manzoor, M. Atif Qureshi, E. Kidney, et al. A Review on Machine Learning Methods for Customer Churn Prediction and Recommendations for Business Practitioners [J]. *IEEE Access*, 2024, 12: 70434-70463.
- [24] 杜睿山, 黄玉朋, 孟令东, 等. 基于BiLSTM-XGBoost混合模型的储层岩性识别[J]. *计算机系统应用*, 2024, 33(6): 108-116.
- [25] 刘中强, 邹维维. 基于采样技术和LightGBM的客户用电异常检测模型 [J]. *计算机系统应用*, 2021, 30(9): 232-236.



- [26] 白子诚, 周艳玲, 张龔. GM-FastText 多通道词向量短文本分类模型 [J]. 计算机系统应用, 2022, 31(9): 403-408.
- [27] 负恺, 贾荣浩, 魏国辉, 等. 基于 CNN 与 Transformer 混合模型的肺炎辅助诊断 [J]. 计算机系统应用, 2025, 34(2): 216-224.
- [28] M. D. Gabhane, A. Suriya, and S. B. Kishor. Churn Prediction in Telecommunication Business using CNN and ANN[J]. Journal of Positive School Psychology, 2022, 6(4): 4672-4680.
- [29] A. Rai. Explainable AI: from black box to glass box.[J]. Journal of the Academy of Marketing Science, 2020, 48(1): 137-141.
- [30] M. Fisher, M. Robb, D. M. Wolf, et al. Implementing a Standardized Discussion Forum Rubric Across 3 Online Nursing Programs [J]. Nurse Educator, 2019, 44(6): 291-292.
- [31] Jiang F, Li Z, Wang Y, et al. A Comprehensive Survey of Large AI Models for Future Communications: Foundations, Applications and Challenges[J/OL]. ArXiv, 2025[2025-03-01]. <http://arxiv.org/abs/2505.03556>.
- [32] Wan H, Zhang L, Zhao Y, et al. Automating Modelica Module Generation Using Large Language Models: A Case Study on Building Control Description Language [J/OL]. ArXiv, 2025 [2025-03-01]. <http://arxiv.org/abs/2509.14623>.
- [33] 靳东明, 金芝, 陈小红, 等. ChatModeler: 基于大语言模型的人机协作迭代式需求获取和建模方法 [J]. 计算机研究与发展, 2024, 61(2): 338-350.