

# 基于信任机制和 TabNet 的京津冀地区共享民宿预定量 影响因素实证研究

## 摘 要

共享经济作为一种通过对有限的资源进行整合,实现各方利益最大化的新型商业模式,近年来在国内外迅速推广开来。同时,随着人们旅游消费需求的日益提高,消费者对于其住宿环境也提出了更高的要求。在此背景下,民宿作为一种共享经济体制下的新型住宿模式应运而生,根据自身需求通过互联网平台搜索进行民宿房间预定的形式也受到了越来越多人的青睐。

与多数线上交易一样,网络带来的虚拟性特征让消费者在预定民宿时心理受到阻碍,其中最重要的影响因素就是信任,因此构建信任机制下的民宿预定模式是实现民宿共享经济发展的核心。为进一步测度这种信任模式对消费者预定行为的影响,本文选择国内行业领先的木鸟民宿作为研究对象,利用 Python 网络爬虫技术获取京津冀地区相关的民宿信息。对于所获取的数据,采用 LDA 主题模型、词云图和共线网络分析获取房东描述和房客评价的真实需求,结合信任机制进行民宿预定量的指标体系构建。最后,针对预处理后的数据进行分析,通过对比不同参数下多种模型在 MSE 等指标上的表现,本文选择 TabNet 模型对于民宿预定量的影响因素进行实证研究,研究结果显示:在特征信任层面,房源特征是消费者在预定共享民宿时考虑程度最低的因素;在认知信任中,房主接单率、房主在线回复率、房源综合得分对共享民宿预定量有正向影响。情感信任所属特征的重要程度最高,其中房源好评率、房主好评率以及房客评价会对共享民宿预定量起到显著提升作用。

本研究具有一定的实践意义和创新价值,采用深度学习模型进行问题研究,打破了传统统计分析方法和数据收集模式的限制。其次,本文结合文本挖掘方法进行指标选取,使得构建的指标体系更具有科学性。根据研究结果,对平台及房主提出相应的建议:平台应完善共享民宿平台的评论及评分机制,全方位提高房源及房主信息的展示功能。房主应该主动提高服务质量,同时对房源进行合理定价,让消费者得到物有所值的服务与产品。

**关键词:** 民宿; 信任机制; 影响因素; 文本挖掘; TabNet

基于信任机制和 TabNet 的京津冀地区共享民宿预定量  
影响因素实证研究

目录

一、 引言 .....1

    (一) 研究背景与意义 .....1

    (二) 文献综述 .....1

    (三) 研究内容与创新点 .....3

        1. 研究内容 .....3

        2. 创新点 .....4

二、 相关概念和理论基础 .....4

    (一) 相关概念和理论 .....4

        1. 民宿定义 .....4

        2. 信任机制 .....5

    (二) 主要模型介绍 .....6

        1. Bert 模型 .....6

        2. TabNet 神经网络模型 .....6

三、 基于文本挖掘与信任机制的指标体系构建 .....7

    (一) 研究对象 .....7

    (二) 基于文本挖掘的房主描述与房客需求分析 .....8

        1. 词云分析 .....8

        2. 主题分析 .....9

        3. 共线网络分析 .....11

    (三) 影响共享民宿预定量指标选取 .....13

        1. 基于特征的信任指标选取 .....13

        2. 基于认知的信任指标选取 .....14

        3. 基于情感的信任指标选取 .....14

四、 京津冀地区共享民宿预定量影响因素实证分析 .....14

    (一) 民宿样本的描述性统计 .....14

        1. 房源特征分析 .....14

        2. 房主特征分析 .....15

    (二) 民宿样本的数据预处理 .....15

        1. 文本数据量化 .....15

        2. 缺失值处理 .....16

    (三) 模型实验环境及评价指标 .....16

    (四) 实验模型选择 .....17

    (五) TabNet 模型训练 .....19

        1. 网络搭建与超参数选取 .....19

        2. 实验结果 .....21

        3. 模型结果解释 .....22

五、 结论与建议 .....	24
(一) 结论 .....	24
1. 基于特征信任 .....	24
2. 基于认知信任 .....	25
3. 基于情感信任 .....	25
(二) 建议 .....	25
参考文献 .....	27

## 表格清单

表 1	房东描述主题提取结果 .....	11
表 2	指标说明 .....	13
表 3	房主回复率/接单率情况 .....	15
表 4	六种模型及其实验参数表 .....	17
表 5	TabNet 超参数设置表 .....	20
表 6	TabNet-P 在测试集上表现情况表 .....	22
表 7	各个维度上特征重要性平均值表 .....	24

## 插图清单

图 1	技术路线图 .....	4
图 2	信任整合模型 .....	5
图 3	Bert 情感分析流程图 .....	6
图 4	TabNet 中的 Feature transformer 结构 .....	7
图 5	TabNet 中 Attentive transformer 结构 .....	7
图 6	房东描述与房客评价文本分析框架图 .....	8
图 7	房客评价词云图 .....	9
图 8	房东描述在不同主题下困惑度 .....	10
图 9	九个主题的可视化 .....	10
图 10	房东描述共线网络图谱 .....	12
图 11	房客评价共线网络图谱 .....	12
图 12	房源价格描述分析图 .....	15
图 13	六种统计模型在 MSE、RMSE 和 MAE 上的表现对比图 .....	18
图 14	六种统计模型在 R2 上的表现对比图 .....	19
图 15	TabNet 网络用于预测共享民宿预定量示意图 .....	20
图 16	两种 TabNet 模型学习曲线及性能比较图 .....	21
图 17	TabNet-P 的 Loss、RMSE 及 MAE 学习曲线 .....	21
图 18	TabNet-P 预测值和真实值比较图 .....	22
图 19	局部特征重要性掩码 Mask[i]（在第 i 决策步中选择的特征） .....	23
图 20	全局特征重要性 .....	23

# 一、引言

## （一）研究背景与意义

自十八大提出新发展理念以来，“互联网+共享”模式催生了一系列共享经济产业。旅游行业消费不断升级，个性化、多样化的住宿需求也推动了共享民宿的产生与发展。房东将闲置房源短租，避免了资源浪费并获得了报酬，相较传统酒店，共享民宿性价比高、社交性强，满足了游客住宿多样化需求。2020年7月，发改委印发的《关于支持新业态新模式健康发展 激活消费市场带动扩大就业的意见》明确提出鼓励共享住宿等产品智能化升级和商业模式创新，发展生活消费新方式。

共享民宿以 C2C 模式为主，即房客与房东个人交易，平台只作为中介提供信息。这种模式耗费资源较少，但也存在一系列问题，如买卖双方信息不对称。互联网的开放性、不可见性，民宿行业的特殊性使该问题潜在风险格外突出<sup>[1,2]</sup>。由于共享民宿提供体验型服务，在未实际消费之前无法有准确的质量感知；用户面对的是完全陌生的房东和其他用户，人身和财产安全保护有很大程度的不确定性，这两点说明房客在选择时会倾向于信任感更强的房东。在疫情防控 and 监管趋严的情况下，信任问题对发展放缓的共享民宿格外重要。基于此，本文从信任视角出发，对共享民宿预定量影响因素进行实证研究，对共享平台和房东提出相关建议。

## （二）文献综述

国内外对共享民宿的研究，主要集中于消费者行为和购买意愿影响因素。赵建欣（2017）<sup>[3]</sup>利用 Airbnb 平台数据，以房源收藏量表示预定意图进行回归分析，发现评论数量、价格和总评分对预定意图有正向作用。Chen（2018）<sup>[4]</sup>对 Airbn

b 用户设计调查问卷，进行方差分析和路径分析，得出房客的感知价值和满意度对购买意愿起决定作用。与相关理论结合，梁晓蓓（2018）<sup>[5]</sup>基于平台 Airbnb 的爬虫数据，从信号理论出发，通过泊松回归探究发现商家的售后互动、特色介绍、回复时间正向影响房源网络口碑，从而积极影响预定量。卢东（2021）<sup>[6]</sup>通过扎根理论的定性研究，解释了共享民宿的产品服务质量、社会影响、安全利益、个人知识等会刺激房客感知有用心理，从而决定消费意愿。在文本挖掘方面，Liu（2021）<sup>[7]</sup>对小猪民宿在线评论提取主题，计算情感得分，与计量经济模型相结合说明了房客体验、感知价值与房客忠诚之间存在正相关，且在影响房客感知价值方面，环境与人际互动同样重要。

从信任层面来看，McKnight 和 Chervany（2001）<sup>[8]</sup>提出了一种跨学科类型学，定义了信任倾向、基于制度的信任、信任信念和信任意图四个概念层面的信任结构，并将信任结构与电子商务顾客行为相联系，为之后信任机制在共享民宿中的研究奠定了基础。谢雪梅（2016）<sup>[1]</sup>通过相关分析和路径分析，基于传统的信任理论，研究了房东个人特征对信任形成机制的影响。贺明华（2019）<sup>[9]</sup>从消费者感知视角出发，通过问卷调查发现第三方信用评分显著影响消费者持续共享意愿，且消费者信任在两者之间发挥中介作用。Yang（2019）<sup>[10]</sup>对 Airbnb 用户在线调查，通过最小二乘法结构方程模型分析，发现基于认知的信任机制比基于情感的信任机制更有效。闫励（2020）<sup>[11]</sup>通过回归树模型发现总评论数、房源位置、差评率、相对价格是影响房客信任最主要的因素，且不同地区影响程度不同。徐峰（2021）<sup>[12]</sup>通过认知、情感和制度三种机制建立信任进行回归实证分析，指出民宿描述长度、在线评论数量等都对预定量有积极影响，房东经营民宿数量则负向影响预定量。

综上所述，国内外学者对共享民宿预定量影响因素研究中，与信任机制的结合较为广泛，大部分学者对评论、房源和房东特征展开研究，并证明了信任在影

响预定量中的重要作用。但目前关于共享民宿购买意愿的研究，数据多以问卷调查获取为主，这可能无法完全反映房客实际体验。评论方面多关注评论数量和分数，忽略了在线评论文本内容和情感信息，且研究方法多为回归、结构方程模型等传统统计方法。本文在以往研究基础上，充分挖掘文本内容及情感信息，通过深度学习模型 TabNet 对共享民宿预定量影响因素进行研究，并与其他模型进行比较。

(三) 研究内容与创新点

1. 研究内容

本文运用共享民宿平台真实数据，基于特征、认知与情感的信任机制，同时充分挖掘文本数据以支持指标选取，通过对比选取 TabNet 模型探究共享民宿预定的影响因素。根据实证研究结果，对民宿房东增强房客的信任感、共享平台改进管理提出相应建议。技术路线图如图 1。

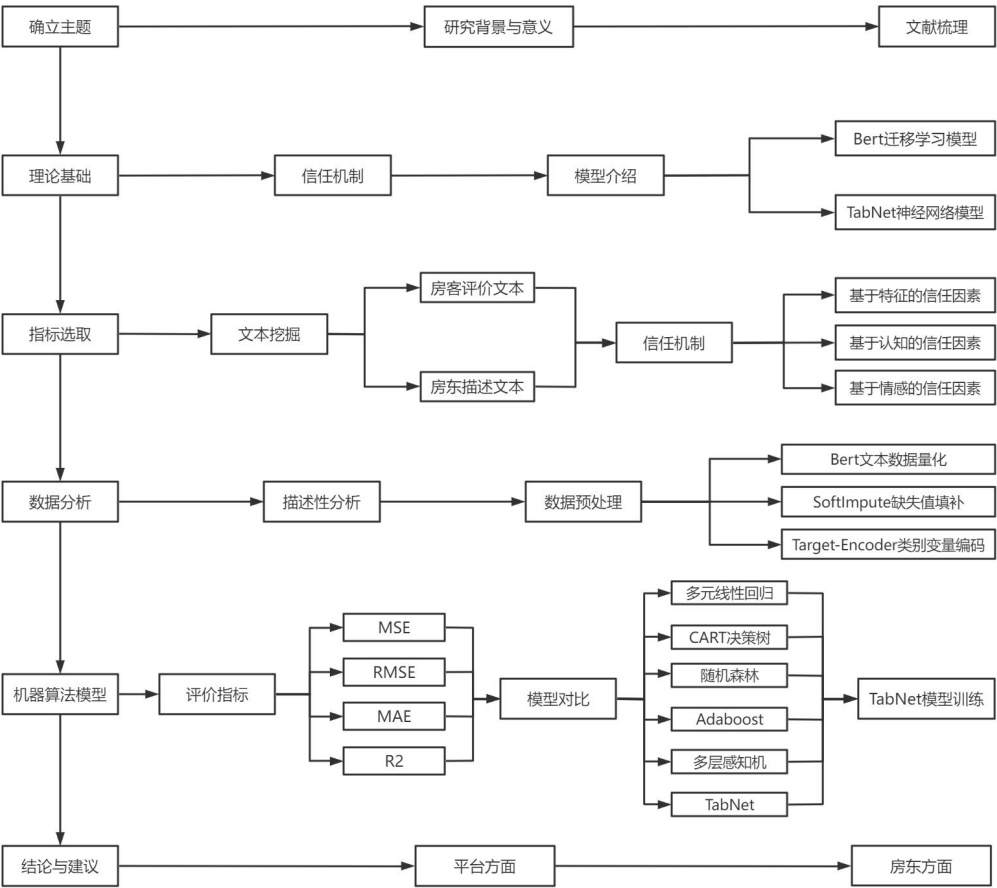




图 1 技术路线图

## 2. 创新点

本文可能的创新点有以下两点：

①从研究方法与模型选择来看，以往的研究大多采用问卷调查获取数据，收集的数据受被调查者影响很大，带有主观色彩；实证研究以相关分析和回归分析为主，模型对数据分布有一定要求，拟合效果可能较差。本文基于平台真实数据，通过深度学习模型 TabNet 探究共享民宿预定量的影响因素，此模型在处理表格数据上具有显著优势，对数据的解释性更强，为该方面研究提供了新的思路。

②从研究视角来看，现有研究指标多以分类变量为主，对评价的研究也仅限于数量和等级，缺少对评论文本的分析。因此，本文通过词云分析、LDA 模型、共线网络对评论内容与房东描述内容进行挖掘，了解房东提供信息主题以及房客的真实需求，并结合信任机制选择指标，解决了指标体系单一问题，构建的指标体系更具科学性与说服力。

## 二、相关概念和理论基础

### （一）相关概念和理论

#### 1. 民宿定义

“民宿”一词起源于国外，早期它是由一些登山、滑雪、游泳等爱好者短期租借民居而衍生并发展起来的一种房屋租赁的使用理念<sup>[13]</sup>。2017 年，国家文化和旅游部明确指出民宿是“利用当地闲置资源，由房屋主人参与接待，为游客其提供体验当地自然、文化与生产生活方式的小型住宿设施。”现阶段我国民宿的经营模式主要是借助互联网技术建立的预定服务平台，消费者可以通过网络查阅民宿房间信息，并通过线上支付的方式来进行预定费用的结算。

## 2. 信任机制

信任的概念源自心理学领域，信任表现出的是一种稳定的心理状态，以及对于其他个体行为或者言语等可信赖性的期望。关于信任机制的研究主要由信任源理论和信任传递理论组成。

### ①信任源理论

对于消费者来说，信任产生的最关键的因素就是信任源，信任源代表的是施信方对信任方自身表现出来的某种特质或属性的感知。在信任源理论涉及的模型研究中，目前认可度最高的是学者 Mayer<sup>[14]</sup>提出的信任整合模型，模型通过将善意、能力和诚实作为感知信任的三个维度，指出信任倾向对于信任的适用情景，同时引入风险指标进行测度，最后观察在一定时间内信任整合模型的预测结果，具体流程如图 2 所示：

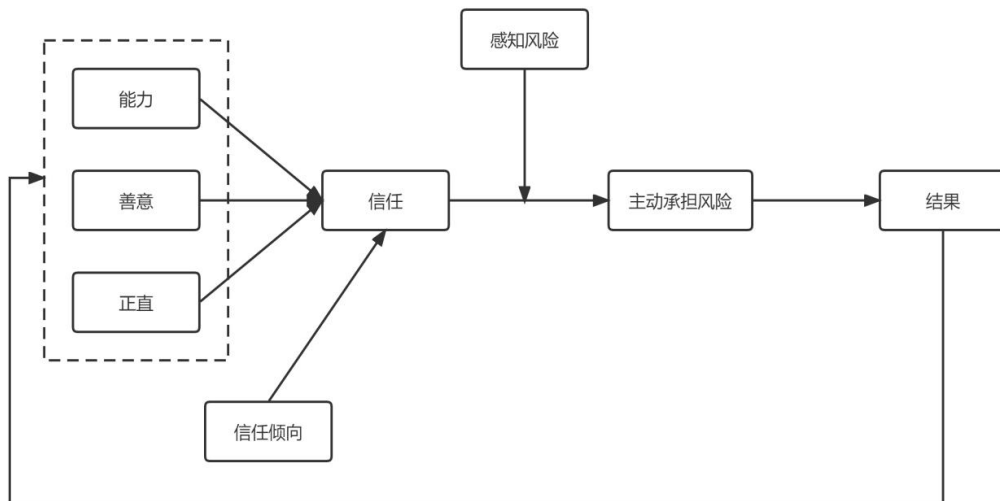


图 2 信任整合模型

### ②信任传递理论

当信任主体对于客体的信任不是通过直接信息进行判断，而是通过去借鉴与信任客体有关的语言、文字以及接触的人等，获得对信任对象的了解，从而进行是否信任的判断<sup>[15]</sup>，此为信任传递。学者吕婷、李君佚<sup>[16]</sup>从社交媒体的角度研究了顾客和商家信任感之间的关系，认为潜在顾客对互联网环境的信任感知会传递到对所购买产品的商家的信任。

## （二）主要模型介绍

### 1. Bert 模型

Bert<sup>[17-18]</sup>是 Google 团队于 2018 年提出的一种语言模型，模型采取双向 Transformer 编码的形式，通过自监督的学习方法，考虑文章的上下结构来调整参数，使得输入的单词学习到高质量的特征表示。Bert 模型在对文本数据进行处理时，首先是要通过构建词库将文本进行分词，再利用词库将切分好的 Token 映射为对应的 ID，将 ID 以 Input Embedding 的形式输入到模型进行预测，如果预测结果中可能有一些错误信息，后续可以利用 Post-Processing 进行改进。

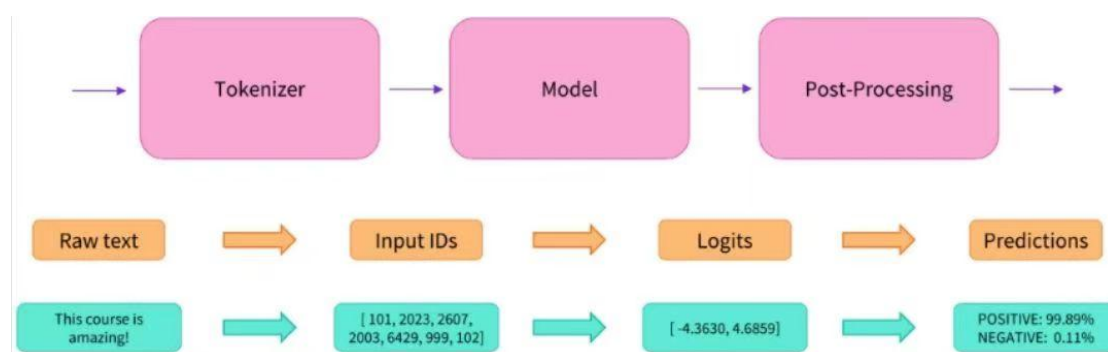


图 3 Bert 情感分析流程图

### 2. TabNet 神经网络模型

TabNet<sup>[19]</sup>是 2019 年 Google 提出的一种用于表格数据的深度学习网络结构，网络采用顺序多步框架，自带特征选择和可解释性于一体。TabNet 中比较关键的是 Feature transformer 层和 Attentive transformer 层，其中 Feature transformer 层采用了两种不同的模块：所有时间步共享 Shared across decision steps 和只影响单步的 Decision step dependent。具体结构如图 4 所示，Feature transformer 先处理共用部分，又接续处理了单步相关模块，保证了模型的稳定性。

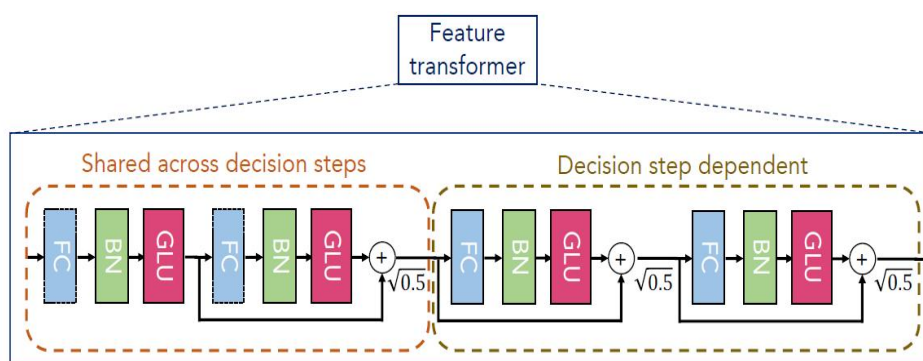


图 4 TabNet 中的 Feature transformer 结构

Attentive transformer 主要作用是在学习时保证权重系数的稀疏性。其具体实现方法是在每一步根据输入特征，得到当前步的输出向量，最终将多步输出向量累加之后，经过全连接层，归一化层，和 Sparsemax 映射出最后的输出 Mask 用于加权处理，通过将原始特征与权重系数 Mask 相乘完成特征选择。

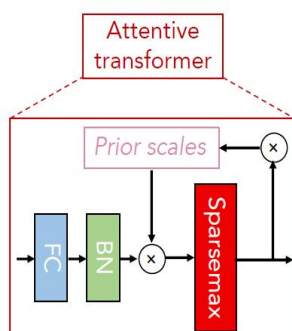


图 5 TabNet 中 Attentive transformer 结构

### 三、基于文本挖掘与信任机制的指标体系构建

#### （一）研究对象

本文综合考虑了民宿预定平台的发展情况和影响力，选取国内发展较为成熟，具有高代表性的木鸟民宿作为本文数据采集的平台，民宿资源丰富、区域协调发展的京津冀城市群作为采集地区，可搜寻到民宿范围的县、市、区作为研究对象，利用 Python 抓取平台上显示的房源特征、房客评价等所有与民宿相关的信息。

## （二）基于文本挖掘的房东描述与房客需求分析

目前关于共享民宿预定量的研究多以经验确定法选取指标，且对于房客评论和房东描述等文本数据的分析浮于表面。根据信任相关理论、其他房客的经历以及房东对房源的描述是否符合房客的需求，会影响房客的信任感知，从而影响预定量。本章尝试从房东描述和房客评价两方面进行关键词、主题、语义网络的分析，具体过程如图 6 所示。

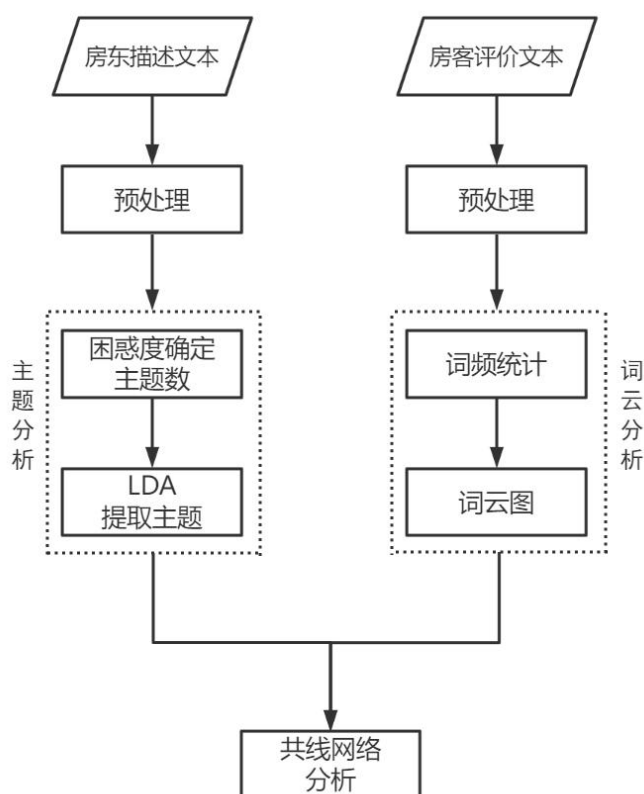


图 6 房东描述与房客评价文本分析框架图

### 1. 词云分析

词云分析是将文本根据词频可视化，以达到突出关键词的作用。本文首先对房客评价进行词频统计，并根据词频对房客评价的关键词绘制词云图，如图 7。

从词云图可以看出，房客评价主要围绕房间和房东两个方面，对房间评价主要体现在干净、设施、齐全、环境、交通、性价比等词汇，对房东评价主要体现在服务周到、热情等词汇。



图 7 房客评价词云图

房客在选择共享民宿时，干净卫生是首要要求，并且基于自身需求，会对周边环境、设施和交通有一定要求，如设有停车场、离地铁较近、离海边较近等。同时，房东的热情回应与周到服务也会增加房客的好感。

## 2. 主题分析

LDA(Latent Dirichlet Allocation)<sup>[20-21]</sup> 是基于三层贝叶斯概率模型的文档主题生成模型，可以用来对给定文档内容构建模型，挖掘文本内容中潜在的主题。本文利用 LDA 模型提取房东描述文本的主题，并得到每条描述对应最大概率的主题。首先，基于困惑度来选择最优的模型主题数，模型生成主题能力越强，困惑度越小，根据图 8，主题数在 8~9 之间比较合适，图 9 显示的是主题数为 9 时主题可视化结果，每个圆圈表示一个主题，圆圈之间离得越远主题相似性越低，主题划分效果越好，最终将主题个数定为 9。

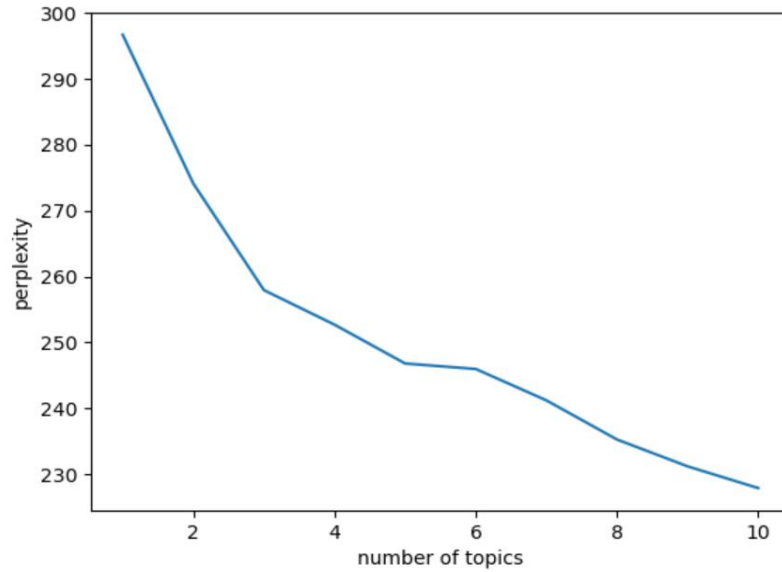


图 8 房东描述在不同主题下困惑度

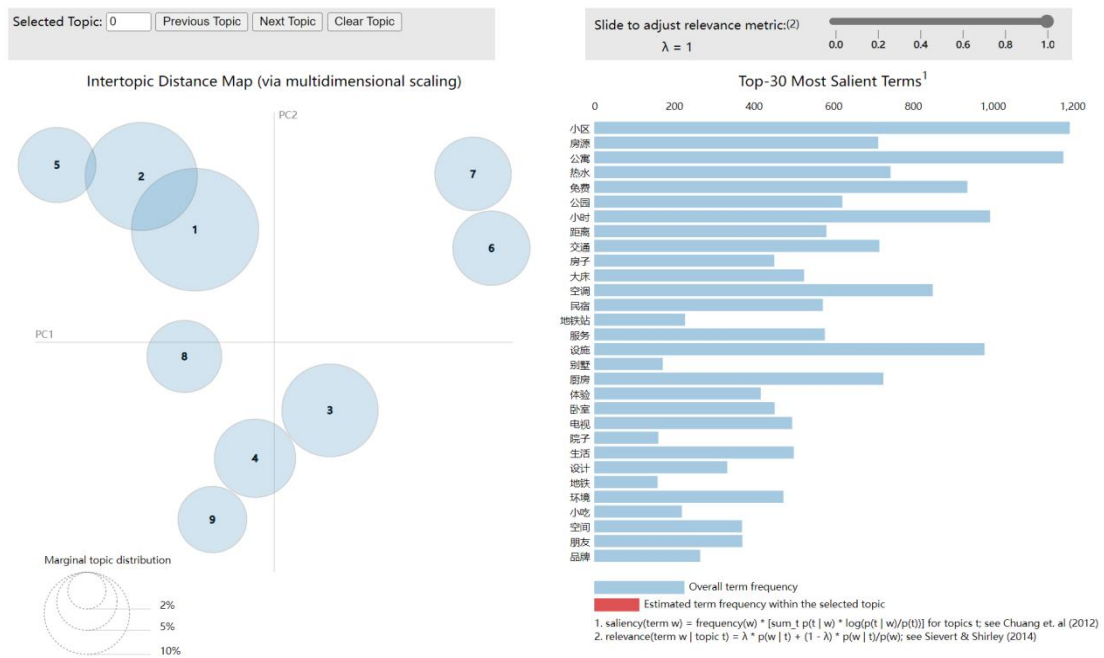


图 9 九个主题的可视化

本文从房东描述文本中提取了九个主题，分别命名为“风格”、“周边”、“房间”、“设备”、“亲友”、“饮食”、“交通”、“类型”和“服务”。表 1 给出了每个主题下包含的 5 个关键词。对每个主题的解释如下：

主题 1：风格。房东对民宿特色描述，如是否带庭院、中式或是日式装修等；

主题 2：周边。民宿周边有哪些便利设施或游乐项目，如火车站、医院等；

主题 3：房间。对民宿布局大小描述，如民宿都包括哪些可使用房间；

主题 4：设备。描述民宿有哪些生活设备，如热水供应、空调冰箱等；

主题 5：亲友。民宿提供给房客怎样的关系体验，如亲子互动、朋友聚会等；

主题 6：饮食。民宿周边有哪些可以吃喝玩乐的场所；

主题 7：交通。民宿地理位置、交通便利情况，如是否靠近车站、地铁站等；

主题 8：类型。民宿户型和入住情况，如小区住宅、拎包入住等；

主题 9：服务。房东可以为房客提供哪些特色服务。

表 1 房东描述主题提取结果

主题	关键词
主题 1	院子 风格 庭院 开间 榻榻米
主题 2	火车站 高铁 广场 医院 公园
主题 3	房间 厨房 卫生间 客厅 卧室
主题 4	热水 空调 电视 洗衣机 冰箱
主题 5	生活 朋友 空间 儿童 亲子
主题 6	小吃 市场 夜市 酒吧 饭店
主题 7	地铁 号线 公交车站 步行 距离
主题 8	小区 房源 户型 拎包 社区
主题 9	服务 行李 特色 电梯 专业

### 3. 共线网络分析

共线网络分析是指在文本中提取某些关键词，利用其之间的相似度构建关系图谱，以揭示文本的内容关联及隐藏信息。图 10 和图 11 分别是根据房东描述与房客评价构建的共线网络图谱，图中每个节点代表一个关键词，节点越大表示该关键词越重要；节点之间连线代表关键词的关联程度，连线越粗关联性越强。

由图 10 可以看出，高中心度词汇主要有小区、热水、用品、冰箱、空调、电视、免费、交通、环境、生活等，与 LDA 主题中关键词基本一致。针对房间的描述，主要涉及方面有设施、环境、户型、空间布局等，与设施关联较强的有小时、房间、空调、洗衣机、热水；与环境关联较强的有小区、公寓等；房间与用品、小区、小时、空调、生活、设施、交通等关联较强。房东可以在描述栏保障基本设施、提高民宿生活质量等，减少房客疑虑从而增强信任感。



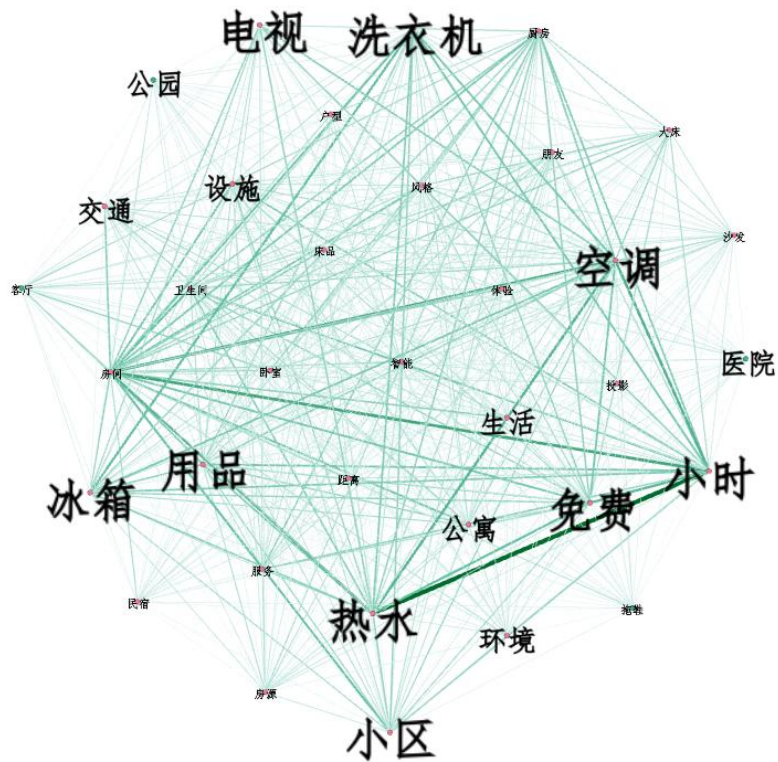


图 10 房东描述共线网络图谱

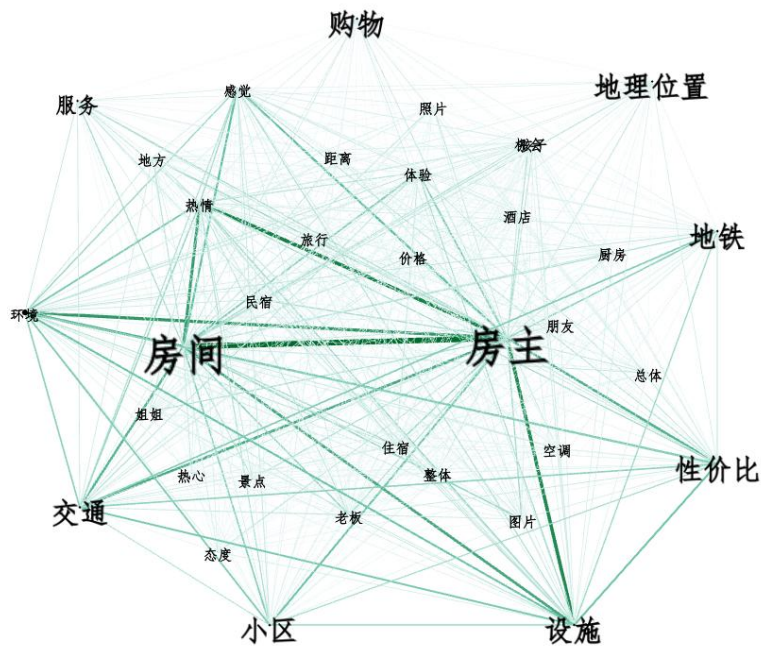


图 11 房客评价共线网络图谱

由图 11 可以看出，高中心度词汇主要有房间、房主、性价比、地理位置、交通、设施、购物、服务等。与房主关联性较强的是服务、热情、态度、感觉、设施等，房主可以考虑热情积极回应房客问题，真诚待人，提高已消费房客体验感，从而改善评价环境，增强房客在首次选择及再次消费的信任。与房间关联性

较强的是环境、性价比、地理位置、交通、图片等，房主除改善基本设施外，也要考虑价格设定是否合理，借助房源真实图片从视觉上增强无法即时消费房客的信任感知。

### （三）影响共享民宿预定量指标选取

根据房东描述的 LDA 主题分析、房客评价的词云分析，以及共线网络分析，结合信任机制的相关理论，将指标分为三个信任维度：基于特征、基于认知和基于情感。表 2 是最终确定的指标汇总及解释。

表 2 指标说明

信任因素	指标名称	指标描述
基于特征的信任	价格	民宿每日房租
	房源类型	民宿是小区住宅或公寓
	出租类型	整租或单间，决定房客是否合租
	可住人数	民宿可容纳最大人数
	面积	民宿总占地面积
	省份	民宿所在省份
	城区	民宿所在区县
基于认知的信任	描述信息量	房东描述文本所含实词数
	房源评论数	民宿评论总数
	房源综合得分	民宿整体得分
	房主回复率	房东回复数量占总咨询量的比例
	房主接单率	房主订单接受率
	其他房源数	房东拥有其他房源数
	其他房源平均分	房东其他房源的平均得分
基于情感的信任	情感得分	房客评价计算的情感得分
	房主好评率	房东好评数占全部评论数量比例
	房源好评率	房源好评数占全部评论数量比例

#### 1. 基于特征的信任指标选取

根据传统信任研究，影响共享民宿信任的因素可分为共享主体个人特征及产品特征<sup>[1]</sup>。从文本分析得出，价格、环境、交通是民宿产品比较重要的特征；不同的关系体验可能对整租单间和容纳人数有不同的需求，民宿是否有庭院、厨房等体现了民宿面积和布局，房源类型也是影响房客选择重要因素。

## 2. 基于认知的信任指标选取

基于认知的信任是指房客理性思考房东和房源是否具有可靠性<sup>[12]</sup>。房客在选择民宿时会首先考虑自身需求，体现在对房东描述文本中信息质量的观察。房客对房东能力感知也会影响认知，接单率、回复率、其他房源数等可以客观地体现房东经营能力。

## 3. 基于情感的信任指标选取

基于情感的信任是房客与房东互动中感受到对方关怀而产生的信心<sup>[22]</sup>，根据信任传递理论，共享民宿的房客主要是通过已消费房客的评价来间接判断房东关怀程度。根据评价词云图，热情、服务周到等词汇体现房东的诚意、善良，表示已消费房客的好评。评价文本量化的情感得分可以直观体现房客情感。

# 四、京津冀地区共享民宿预定量影响因素实证分析

## （一）民宿样本的描述性统计

### 1. 房源特征分析

由房源价格描述分析图 12 可知，房源定价主要集中在 100-199 元以及 200-399 元，由此可推断出该价格区间符合大部分人对于民宿愿意支付的价格水平。定价为 100 元以下的房源消费人数较少，仅为 122 人，可能是因为本身价格偏低的房源住宿环境要差，多数人不太愿意牺牲生活品质来换取价格上的获利，进而导致该价格区间的房源预定量偏低。

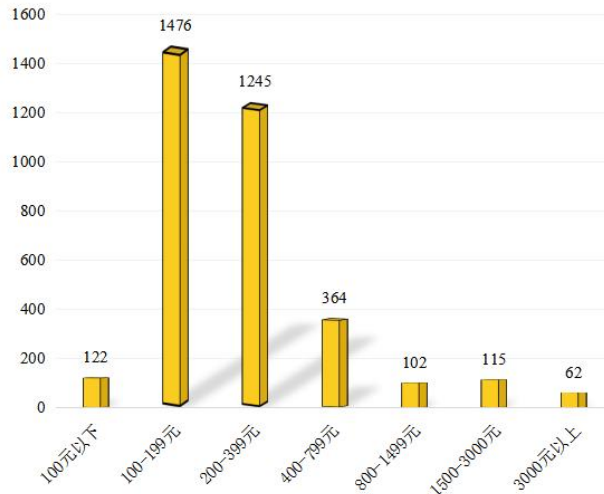


图 12 房源价格描述分析图

## 2. 房主特征分析

房主的回复率和接单率可以反应房主对于房源管理的认真程度,由表 3 可知,80%以下的回复率和接单率情况较少,95%以上的情况较多,样本差异性大。数据中的房主平均回复率超过 95%,接单率的均值也高达 93.46%,民宿房主普遍回复较为积极,接单成功率较高,整体服务态度良好。

表 3 房主回复率/接单率情况

房主回复率			房主接单率	
回复率区间	数量	占比情况	接单率数量	占比情况
70-80%	285	8.18%	0	0
80-90%	56	1.61%	889	25.50%
90%-95%	86	2.47%	416	11.94%
95%以上	3059	87.74%	2181	62.56%
平均回复率	95.33%		93.46%	

## (二) 民宿样本的数据预处理

数据预处理是在正式开始分析前,需要对数据进行审查和校验工作,以达到提高数据的质量、适应模型分析方法的目的。预处理工作主要分为以下阶段:

### 1. 文本数据量化

在本文的民宿样本数据中,房源描述列和房客评价列都是以文字的形式进行

展示。对于房源描述列来说，房主描述的详尽程度包括了影响房客判断的价值信息，可以采用计算字符长度的形式对其进行数值型转化。民宿评价列较为特殊，房客在评论时能依照自身的住宿体验表达出自己的情感倾向，因此本文在对这类数据进行量化处理时，利用 Bert 模型对房客的评价内容进行编码转化后，结合注意力头可视化机制对评价列进行情感打分，并按照分值的高低设定界限进行分类，正确反映其所蕴含的情感倾向。

## 2. 缺失值处理

插补法主要是利用特征本身的统计规律对数据进行填充，或利用模型来估算结果。对于抓取到的民宿数据中可能因房东描述或房客相关信息未填写而造成的样本信息缺失，本文主要采用 SoftImpute<sup>[23-24]</sup>算法来处理这些缺失值，该算法使用迭代的软阈值 SVD (Singular Value Decomposition) 进行估算，通过核范数正则化将低秩矩阵近似拟合到有缺失值的矩阵上，最终利用生成的结果来进行填补。

## (三) 模型实验环境及评价指标

由于机器学习和深度学习模型训练中涉及到对初始民宿数据集的划分，本文将预处理完毕的数据按照 8:1:1 的比例划分为训练集、开发集以及测试集，其中训练集包括 2788 个民宿样本，开发集和训练集分别包括 349 个民宿样本。同时，为保证实证结果的科学性及实验代码的可复现性，各实验均在 CPU 为 i5-8250U 的计算机上进行，实验时采用的环境为 Python3.7。

处理因素分析回归时通常采用的评价指标有均方误差 (MSE)、平均值误差 (MAE)、均方根误差 (RMSE)、R2 决定系数 (R2)、平均绝对百分比误差 (MAPE) 以及对称平均绝对百分比误差 (SMAPE) 等。但鉴于本文所用深度学习模型损失函数选取的问题以及评价指标的普适性，决定采用 MSE、RMSE、MAE 和 R2 作为本文研究的评价指标，其计算公式如下所示：

$$\begin{aligned}
MSE &= \frac{1}{n} \sum_{i=1}^n (ypred_i - ytrue_i)^2 \\
RMSE &= \frac{1}{n} \sum_{i=1}^n \sqrt{(ypred_i - ytrue_i)^2} \\
MAE &= \frac{1}{n} \sum_{i=1}^n |ypred_i - ytrue_i| \\
R2 &= 1 - \frac{\sum_{i=1}^n (ytrue_i - ypred_i)^2}{\sum_{i=1}^n (ytrue_i - ymean)^2}
\end{aligned} \tag{1}$$

其中  $n$  代表测试集上的民宿样本总量,  $ypred_i$  代表第  $i$  个民宿样本的近期预  
定量预测值,  $ytrue_i$  代表第  $i$  个民宿样本的近期预定量真实值,  $ymean$  代表测试集  
上所有民宿样本近期预定量的平均值。

#### (四) 实验模型选择

为保证最终实证分析结果的可靠性,下面对共享民宿数据集在不同模型上的  
表现进行分析并选择出最优的模型进行实验。同时,为获取到最佳的实验模型参  
数,本文研究时采用随机网络搜索的方法得到每个模型的实验参数,选择对比的  
模型及其参数如表 4 所示:

表 4 六种模型及其实验参数表

实验模型	参数	参数值
多元线性回归 (MLR)	训练截距	True
CART 决策树 (DT)	分裂标准	MSE
	树最大深度	20
	森林规模	300
随机森林 (RF)	分裂标准	MSE
	树最大深度	20
	弱学习器	DT
Adaboost (ADB)	学习器数量	130
	学习率	0.1

	网络形状	(100,100)
多层感知机（MLP）	批大小	256
	Epoch	200
	优化器	Adam
TabNet	批大小	256
	Epoch	200

经过对 6 种模型的参数择优后，下面考察每种模型在 MSE、RMSE、MAE 以及 R2 四个指标上的表现，其结果如图 14、图 15 所示：

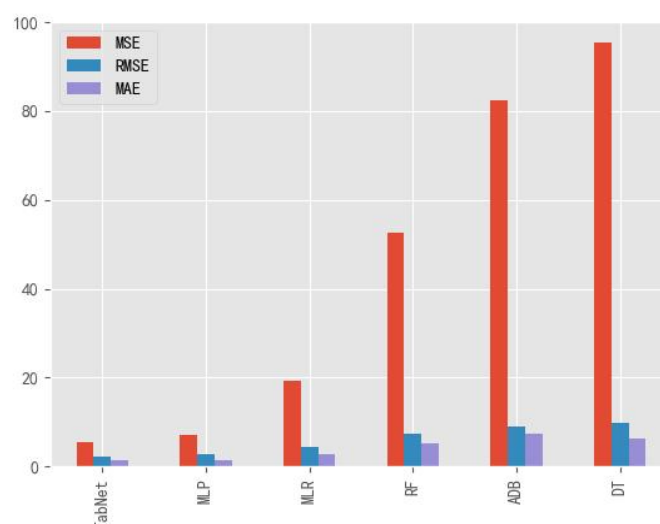


图 13 六种统计模型在 MSE、RMSE 和 MAE 上的表现对比图

由图 13 可以看出，不论观测 MSE 还是 MAE，6 个模型中回归误差最小的都是 TabNet 神经网络。此外，观察上图还能发现这六种模型之间性能差异较大，例如 TabNet 神经网络模型的 MSE 值相较于表现最差的 CART 决策树模型降低了大约 89.962，MAE 降低了大约 4.838，由此可见，在对民宿数据进行建模时模型选择是十分关键的一环。除了衡量预测的民宿预定量和实际的民宿预定量之间的误差外，还可以对其的拟合程度进行分析，实验结果如图 14 所示：

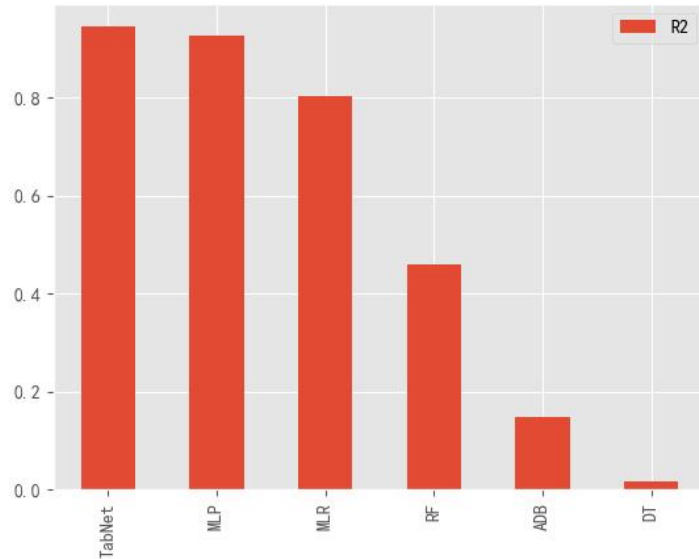


图 14 六种统计模型在 R2 上的表现对比图

其中，TabNet 神经网络输出的近期民宿预定量相较于其他五个模型最接近于预定量的真实值，其拟合程度达到 94.4%，相较于次之的多层感知机模型提高了约 2 个百分点，是最差的 CART 决策树模型的 55.2 倍。综上所述，TabNet 模型具有较高的拟合度和较低的预测误差。因此，本文建立 TabNet 模型对京津冀地区民宿数据进行分析。

## （五）TabNet 模型训练

### 1. 网络搭建与超参数选取

TabNet 神经网络在原理上具有两个优势：其一是此网络模仿的树形结构可以产生一种近似于超平面的边界的决策流形，超平面的边界可以有效地划分数据，使模型能够有效地表示表格数据；其二是此网络具有良好的解释性，可以对京津冀地区共享民宿预定量的预测结果负责。此外，由上一节的论述结果可知，该网络在本实验数据集上表现较为良好，从实践方面论证了将 TabNet 神经网络用于预测共享民宿预定量的可行性。下面将 TabNet 神经网络结构用于本文实验数据集，搭建出的网络模型（示例）如图 15 所示：



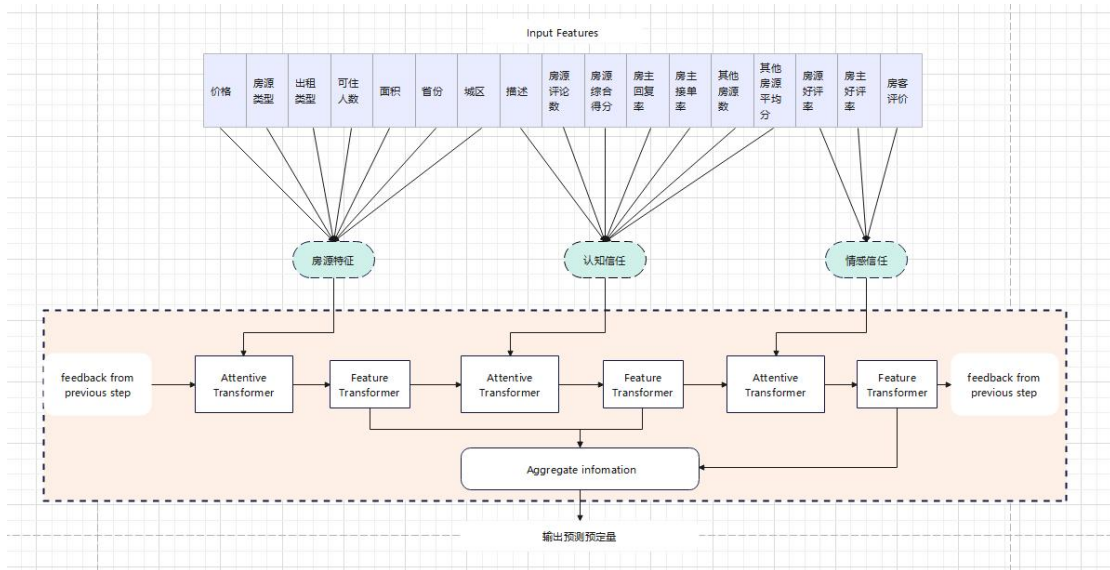


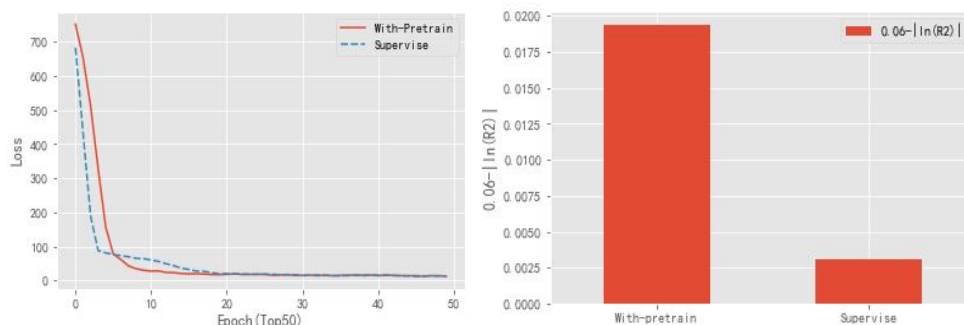
图 15 TabNet 网络用于预测共享民宿预定量示意图

上图是当 TabNet 中决策步长为 3 时网络结构的示意图。决策预测层的宽度、注意力嵌入的宽度和决策步长是决定模网络框架的重要参数,对于大多数数据集,将决策步长设置在从 3 到 10 的区间内且决策预测层的宽度等于注意力嵌入的宽度是合理的<sup>[19]</sup>。此外,经试验降低决策预测层的宽度、注意力嵌入的宽度和决策步长可以有效减少过拟合且不会显著降低精度。综上考虑确定本文 TabNet 的超参数设置如表 5 所示。

表 5 TabNet 超参数设置表

超参数	描述	值
N_d	决策预测层的宽度	8
N_a	注意力嵌入的宽度	8
N_steps	决策步长	3
Lr	学习率	0.02
Optimizer_Fn	优化器	Adam
Batch_Size	批大小	256
Epoch	迭代次数	200

由于 TabNet 网络本身提供了基于自监督学习构造预训练模型 (With-Pretrain) 和监督学习 (Supervise) 两种策略,下面对两种策略分别进行实验分析,选择出最适合本文研究的实验模型。实验结果如图 16 所示:



注：由于二者拟合系数差距不明显，因此对其进行映射处理

图 16 两种 TabNet 模型学习曲线及性能比较图

其中，有监督共享民宿预测模型在第 30 次迭代后开始慢慢收敛，在经过 200 次迭代后其在测试集中预测的 R2 值达到了 0.944；而自监督预训练模型是在第 15 次迭代开始慢慢收敛，同样经过 200 次迭代后其 R2 值可以达到 0.960，说明带有自监督学习预训练的 TabNet（TabNet-P）在保证收敛速度的同时，提高了模型的性能。所以对于研究京津冀共享民宿预定量运用 TabNet-P 模型更为合适。

## 2. 实验结果

为深入探究影响京津冀地区共享民宿预定量各个指标之间的联系以及各个指标对预测结果所产生的贡献，本文利用 TabNet-P 神经网络对京津冀地区共享民宿数据集进行实验分析，实验结果如图 17 所示：

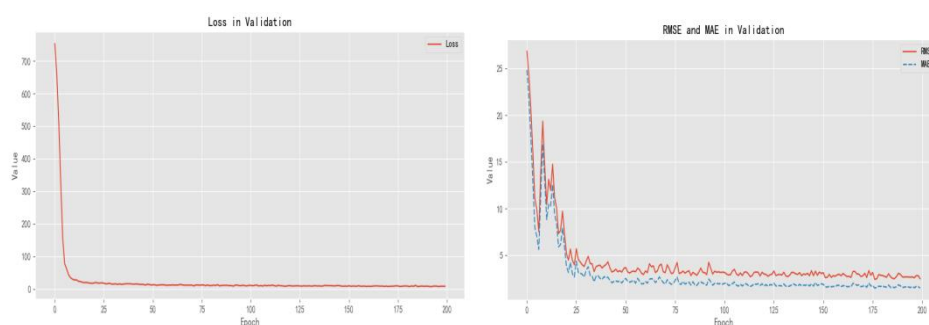


图 17 TabNet-P 的 Loss、RMSE 及 MAE 学习曲线

由图 17 可知，TabNet-P 的 Loss 学习曲线在学习迭代 15 次左右之后趋近于收敛，而 RMSE 和 MAE 曲线则是在学习迭代 100 次后开始趋向于收敛。而模型在学习迭代 200 次后，在开发集上的 Loss、RMSE 和 MAE 分别为 8.06、2.47 以

及 1.49, 利用此时训练完的网络模型在样本容量为 349 的测试集上检测模型性能, 实验结果如表 6 所示:

表 6 TabNet-P 在测试集上表现情况表

观察指标	值
MSE	3.88
RMSE	1.97
MAE	1.17
R2	0.96

表 6 中给出了 TabNet-P 神经网络在 MSE、RMSE、MAE 和 R2 四个指标上的表现。由衡量误差程度的指标可知, 利用该神经网络在对京津冀地区民宿预定量进行预测时其平均误差结果正负不会超过 2 个单位; 由衡量拟合程度指标可知, R2 值达到了 0.96 十分接近于 1, 说明该神经网络的预测值接近于真实值, 图 18 为测试集上 349 个民宿样本真实值和预测值拟合图。

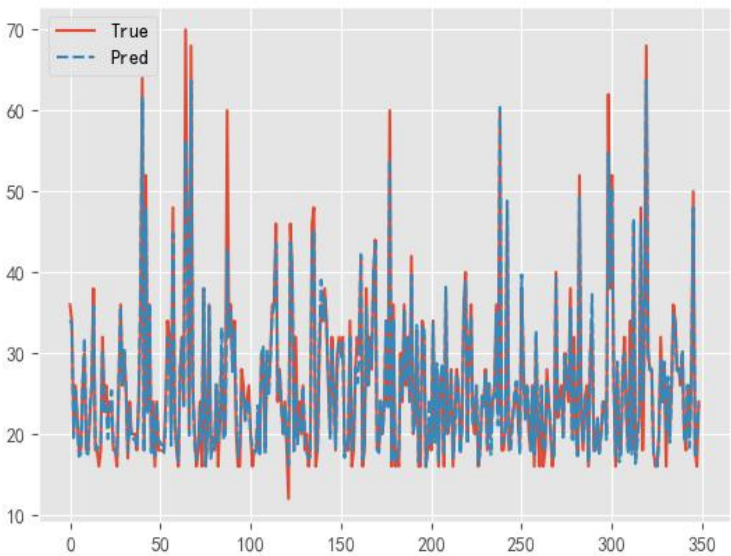
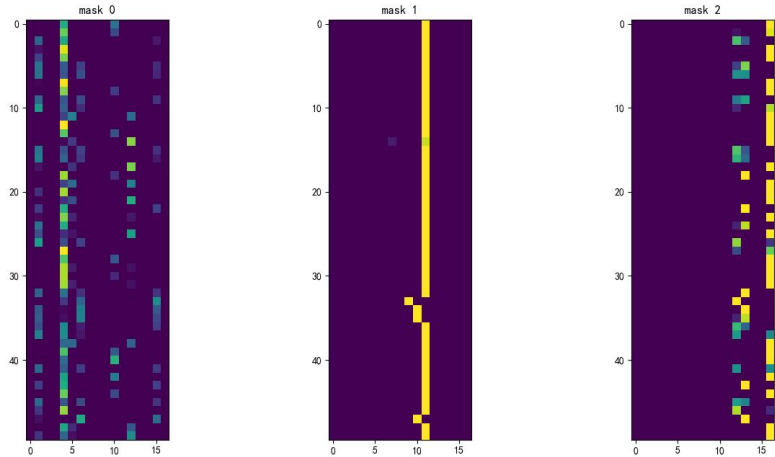


图 18 TabNet-P 预测值和真实值比较图

### 3. 模型结果解释

根据调参后的 TabNet-P 模型进行训练, 得到每个决策步 (Step) 变量的重要程度, 可以看出在第一个决策步内房源类型、可住人数、面积, 省份等特征被赋

予了较大权重，所以模型在本决策步内选择部分变量进行特征计算；在第二个决策步内模型仅仅选择了房客评价、房源综合得分等特征进行计算；之后模型在第三个决策步内认为房主接单率、其他房源数、房源好评率等特征重要，如图 19。



注：颜色越明亮、特征中明亮的样本越多代表该特征越重要

图 19 局部特征重要性掩码 Mask[i]（在第 i 决策步中选择的特征）

将各个决策步中特征重要性进行汇总计算得到全局特征重要性并将其进行从大到小排序得到图 20。在此次 TabNet-P 模型预测中，影响因素的重要性大小排名前十的特征依次为：房主好评率、出租类型、房源评论数、房主回复率、房主接单率、房客评价、房源好评率、房源类型、价格、房源综合得分，说明这些影响因素在共享民宿销售量影响因素中影响较大。面积、可住人数、省份、描述、其他房源平均分、其他房源数和城区对共享民宿销售量的影响随之减少。

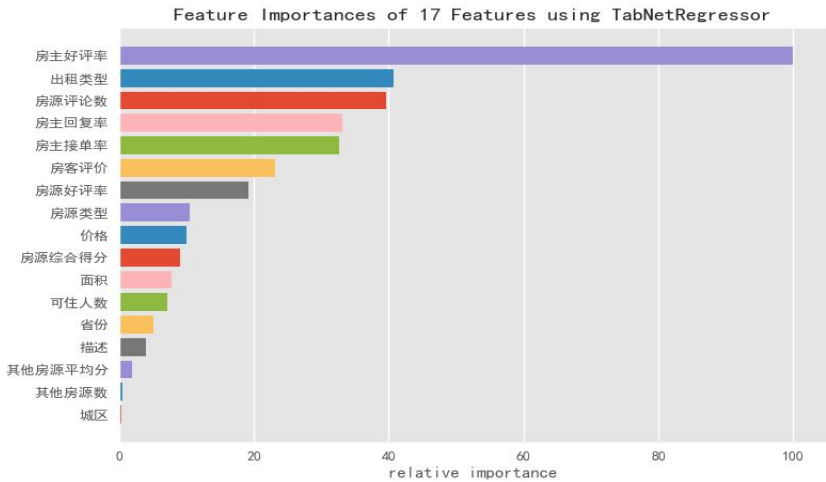


图 20 全局特征重要性

此外，上述的 16 个特征分别是基于房源特征、房客评价以及信任机制等维度进行选取的，所以对基于特征信任、认知信任以及情感信任等维度中具体特征计算平均重要程度更有实际意义，下面对每个维度下所属特征重要性进行平均计算，结果如表 7 所示。

表 7 各个维度上特征重要性平均值表

特征维度	重要性平均值	重要性占比（%）
基于特征的信任	11.243	13.21
基于认知信任	27.959	32.85
基于情感信任	45.899	53.93

其中，情感信任所属因素特征的平均重要程度最高为 0.1377，占据总重要性的 53.93%，说明消费者在面临民宿选择时主要看重情感信任所属因素；其次认知信任所属因素的平均重要程度仅次于情感信任，其值为 0.08388，重要性占比为 32.85%；三个维度中重要性最低的是房源特征因素，其值为 0.03373，该维度重要性占比仅有 13.21%。

## 五、结论与建议

### （一）结论

本文以京津冀地区共享民宿为研究对象，从共享经济的视角下探究了影响该地区共享民宿预定量的主要因素，通过实证分析研究发现如下结论：

#### 1. 基于特征信任

通过实验研究发现，房源特征是消费者在面临预定共享民宿问题时考虑程度最低的一个维度。在此维度下，消费者更看重的是出租类型、房源类型以及价格，出租类型和房源类型决定着消费者入住的房间概况，而价格方面则取决于消费者的可支配收入水平；另一方面，面积、可住人数、城区等特征的重要程度相对较

低，这是因为该地区消费者在选择民宿入住时通常来说会轻装出行，并不会携带过多的大宗行李物品，所以对于面积、可住人数等因素要求相对较低。此外，在以一个城市为中心点的附近城区内共享民宿会展现出同质性的特点，因此对于消费者来说城区特征重要程度也相对较低。

## 2. 基于认知信任

在该维度所属的因素中，代表房间受欢迎程度的房源评论数，代表能力的房主接单率、代表善意的房主在线回复率、代表服务质量的房源综合得分均会对共享民宿预定量起正向显著作用，且民宿的房源评论数对共享民宿预定量影响程度最高。相反，房主对于房间的描述信息量、房主的其他房源数以及其他房源平均分的重要性相对较低，这是因为当消费者在选择民宿时，进入木鸟民宿网站主页直接会看到对于此房源的相关信息，而房主的其他房源信息较为隐蔽，所以导致了消费者对于房主其他房源的信息了解的较少。

## 3. 基于情感信任

采用 TabNet-P 对京津冀地区共享民宿数据进行拟合实验可知，基于情感信任所属特征的重要程度最高，占据了总重要性的 53.93%。在该维度所属的特征中，房源好评率、房主好评率以及房客评价均会对共享民宿预定量起十分显著的作用。这是因为在口碑理论与消费者购买决策理论中，潜在消费者在购物过程中会关注评论信息，从而影响购买意愿，因此评论中所反映的信息十分重要。综上，房主应特别关注评论内容中所反映的各个方面，结合预定量的影响因素做出改善以提高在线民宿预定量。

## （二）建议

基于以上研究结论，本文对于平台和房主提出以下相关建议：

第一，完善共享民宿平台的评论及评分机制。经过本文研究发现，房主好评

率、房源好评率、房客评价和房源综合得分等是影响该地区共享民宿预定量的重要因素。因此，平台可以完善评论以及评分机制，采取相应措施提高消费者对所租用房间进行评论的积极性，评论的字数越多其包含的信息量就越大，而这不仅可以使得其他消费者更全面客观的了解该房源信息，提高其购买意愿，还可以帮助房主了解消费者的真实想法，挖掘消费者的关注点，以帮助房主提升自身服务质量和完善房源设施。

第二，全方位提高房源及房主信息的展示功能。经过本文研究发现，在基于认知信任选取因素的维度中，房主的其他房源数和其他房源平均分等因素的重要性相对较低，主要是因为当消费者进入网络平台选择民宿时，由于房主其他房源信息较为隐蔽，所以导致了消费者对于房主其他房源的信息了解的较少。故平台应将此部分隐蔽信息放置在明显的位置进行展示或对其建立快捷通道使消费者能够快速浏览到房源以及房主的全部信息，进一步可以减少由于信息不对称等原因带来的消费者不信任，最大程度上保证了房源的真实度以及房主的诚信度。

## 参考文献

- [1]谢雪梅, 石娇娇. 共享经济下消费者信任形成机制的实证研究[J]. 技术经济, 2016,35(10):122-127.
- [2]范秋敏. 基于信任理论的在线民宿产品消费者购买意愿研究[D]. 北京交通大学, 2019.
- [3]赵建欣, 朱阁, 宋玲玉. 在线短租平台用户住宿决策影响因素研究[J]. 北京邮电大学学报(社会科学版), 2017,19(05):52-57.
- [4]Chen C C, Chang Y C. What drives purchase intention on Airbnb? Perspectives of consumer reviews, information quality, and media richness[J]. Telematics and Informatics, 2018, 35(5): 1512-1523.
- [5]梁晓蓓, 徐真, 李晶晶. 共享短租平台商家属性对消费者网络口碑的影响研究[J]. 数据分析与知识发现, 2018,2(11):46-53.
- [6]卢东, 曾小桥, 徐国伟. 基于扎根理论的共享住宿选择意愿影响因素及机理研究[J]. 人文地理, 2021,36(02):184-192.DOI:10.13959/j.issn.1003-2398.2021.02.022.
- [7]Liu F, Lai K H, Wu J, et al. Listening to online reviews: a mixed-methods investigation of customer experience in the sharing economy[J]. Decision Support Systems, 2021, 149: 113609.
- [8]McKnight D H, Chervany N L. What trust means in e-commerce customer relationships: An interdisciplinary conceptual typology[J]. International journal of electronic commerce, 2001, 6(2): 35-59.
- [9]贺明华, 陈文北. 共享经济平台信用机制对持续共享意愿的影响——消费者信任的中介作用[J]. 中国流通经济, 2019,33(05):66-80.DOI:10.14089/j.cnki.cn11-3664/f.2019.05.008.



- [10]Yang S B, Lee K, Lee H, et al. In Airbnb we trust: Understanding consumers'trust-attachment building mechanisms in the sharing economy[J]. International Journal of Hospitality Management, 2019, 83: 198-209.
- [11]闫励, 牛新艳. 我国共享短租市场中的信任研究——基于 Airbnb 开源数据的分析[J]. 经济论坛, 2020(12):137-147.
- [12]徐峰, 张新, 梁乙凯, 等. 信任构建机制对共享民宿预订量的影响——基于 Airbnb 的实证研究[J]. 旅游学刊, 2021, 36(12): 127-139.
- [13]吴晓隽, 于兰兰. 民宿的概念厘清、内涵演变与业态发展[J]. 旅游研究, 2018, 10(02): 84-94.
- [14] MAYER R C, DAVIS J H, SCHOORMAN F D. An integrative model of organizational trust[J]. Academy of Management Review, 1995(20):709-734.
- [15]汪黄梅, 丁文辉. 基于信任理论的民宿预定量影响因素研究——以长三角民宿区为例[J]. 福建商学院学报, 2019(01): 37-43. DOI: 10.19473/j.cnki.1008-4940.2019.01.006.
- [16]吕婷, 李君轶, 代黎, 王萌, 杨敏. 电子口碑对乡村旅游行为意向的影响——以西安城市居民为例[J]. 旅游学刊, 2018, 33(02): 48-56.
- [17]Devlin J, Chang M W, Lee K, et al. Bert: Pre-training of Deep Bidirectional Transformers for Language Understanding[J]. 2018.
- [18]赵宏, 傅兆阳, 赵凡. 基于 BERT 和层次化 Attention 的微博情感分析研究[J]. 计算机工程与应用, 2022, 58(05): 156-162.
- [19]Sercan O. Ark, Tomas Pfister. TabNet: Attentive Interpretable Tabular Learning
- [20] Blei D M, Ng A Y, Jordan M L. Latent Dirichlet Allocation[J]. Journal of Machine Learning Research, 2003, 3(1):993-1022.
- [21]肖明, 商慧语, 肖毅, 廖莉莉. 基于 LDA 模型的统计学热门主题挖掘及知识

- 图谱分析[J/OL].华中师范大学学报(自然科学版): 1-11[2022-04-30].
- [22]牛阮霞, 何砚.共享住宿中的信任研究综述——基于房东、房客视角[J].管理现代化,2020,40(03):112-117.DOI:10.19634/j.cnki.11-1403/c.2020.03.029.
- [23] Troyarskaya O, Cantor M , Sherlock C,et al.Missing value estimation methods for DNA microarrays[J]. Bioinformatics ,2001,17(6):520-525.
- [24]王媛, 温阳俊, 王艳萍, 刘汉钦, 马若洵, 吴清太, 张瑾.自然群体多性状表型缺失值预测方法的比较[J].南京农业大学学报, 2022, 45(02): 395-403.