

# 产品订单的数据分析与需求预测

## 摘要

产品订单的需求量预测对公司管理层指定合理的销售计划、目标和策略都是极为重要的。对产品订单需求的准确预测可以帮助企业解决库存积压和资金不足等问题。

针对问题一，我们对需求量历史数据进行了数据分析，获取构建特征工程的思路。首先我们进行了数据清洗，包括识别重复值、异常值、缺失值并进行处理。我们分析了价格、销售区域、销售方式、品类、时间段等基本信息和产品需求量的关系，以探求不同特征对需求量的影响。根据数据分析的结果，我们发现节假日、季节，交易日期是否处于月初月末等因素都会对需求量产生较大的影响。

我们根据前面的数据分析，从三个方向出发进行特征构造：时间戳特征、时序值相关特征和属性变量特征。我们构造的特征包含了从交易事件中提取的时间特征与从真实场景出发构造的特征。在特征构造之后，我们进行了特征选择去除冗余和无效特征，以达到减少模型学习时间、提高模型泛化能力的目的。

针对问题二，使用前面的特征构造，分别按月、周、日三种时间粒度预测商品未来需求量。每种时间粒度都构建了优化了模型结构和超参数的 BiLSTM、BiLSTM-Attention、LightGBM 三类模型，其中天时间粒度多构造了 LightGBM-BiLSTM-Attention 组合模型。我们通过记录比较各模型的误差评价指标值和拟合效果来判断预测精度，发现使用 LightGBM 算法模型按天时间粒度预测精度较高。

**关键词：**BiLSTM；BiLSTM-Attention；LightGBM；LightGBM-BiLSTM-Attention；组合模型

# 目录

1 绪论 .....	1
1.1 问题背景 .....	1
1.2 问题重述 .....	1
1.3 主要工作 .....	2
2 数据预处理 .....	3
2.1 缺失值情况说明与处理.....	3
2.2 异常值情况说明与处理.....	3
2.3 重复值情况与处理.....	3
3. 数据分析 .....	4
3.1 价格对需求量的影响.....	4
3.2 区域对需求量的影响.....	5
3.3 销售方式对需求量的影响.....	5
3.4 品类对需求量的影响.....	7
3.5 时间段对需求量的影响.....	8
3.6 节假日对需求量的影响（假日：1，非假日：0） .....	12
3.7 促销对需求量的影响.....	13
3.8 季节对需求量的影响.....	13
4 特征工程构建 .....	15
4.1 特征构造 .....	15
4.1.1 时间戳衍生特征.....	15
4.1.2 时序值衍生特征.....	18
4.1.3 属性变量.....	19
4.2 特征选择 .....	20
5. 预测算法原理与评价指标.....	22
5.1 BiLSTM 算法原理 .....	22
5.2 BiLSTM-Attention 算法原理.....	23
5.3 LightGBM 算法原理.....	24
5.4 集成学习与组合算法原理.....	25

5.5 预测算法误差评价指标.....	26
6 产品需求量预测 .....	28
6.1 按月时间粒度预测.....	28
6.1.1 数据集处理.....	28
6.1.2 单/多变量 BiLSTM 模型构建与比较 .....	28
6.1.3 多变量 BiLSTM 模型调参 .....	30
6.1.4 多变量 BiLSTM-Attention 模型构建.....	31
6.1.5 LightGBM 模型构建.....	33
6.1.6 预测精度对比与 LightGBM 模型选择 .....	34
6.2 月、周、天时间粒度预测精度对比.....	35
6.2.1 数据集处理.....	35
6.2.2 按周粒度模型预测精度对比.....	35
6.2.3 按天粒度模型预测精度对比.....	36
6.2.4 选取预测精度最高的粒度——天粒度.....	37
6.3 按天时间粒度的预测模型选择.....	37
6.3.1 LightGBM-BiLSTM-Attention 组合模型.....	37
6.3.2 选取 LightGBM 预测模型.....	38
7 总结 .....	40
参考文献 .....	42

# 1 绪论

## 1.1 问题背景

近年来，随着市场竞争的加剧和消费者需求的不断变化，企业外部环境越来越不确定，复杂多变的外部环境，让企业供应链面临较多难题。需求预测作为企业供应链的第一道防线，重要程度不言而喻，然而需求预测的准确性受到多种因素的影响，其中包括市场竞争、消费者行为、经济环境、政策法规等方面。各种因素的变化会导致市场需求的波动和不确定性，导致预测准确率普遍较低，因此需要更加优秀的算法来解决这个问题。

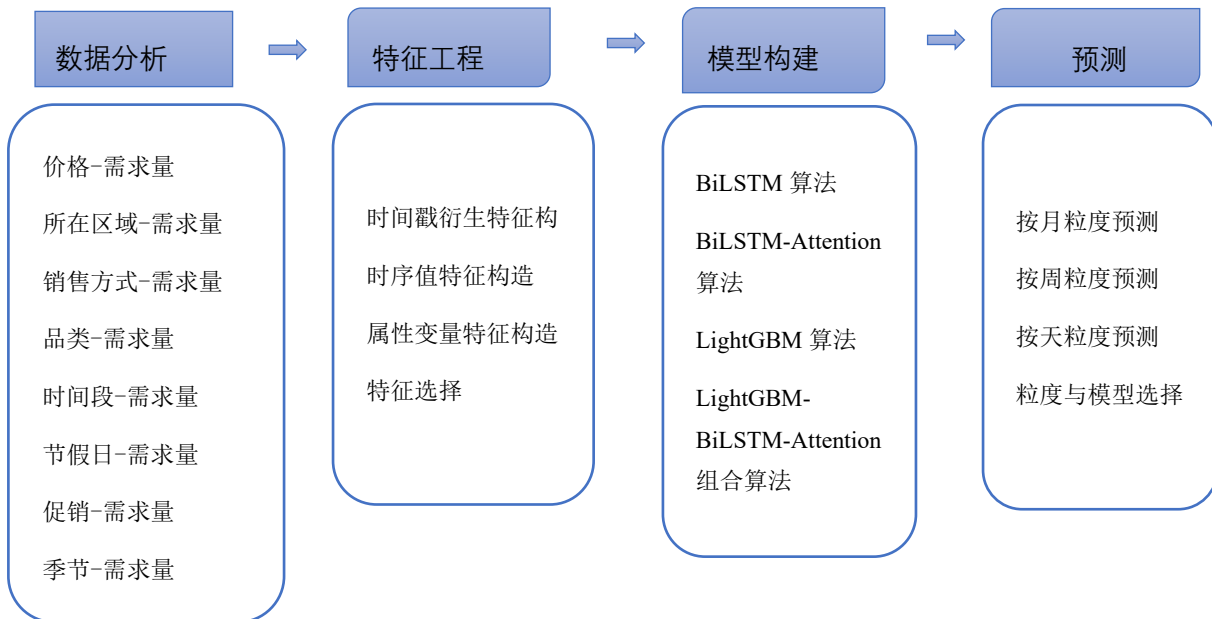
需求预测是基于历史数据和未来的预判得出的有理论依据的结论，有利于公司管理层对未来的销售及运营计划、目标，资金预算做决策参考，准确地预测未来销售额和市场份额，从而制定相应的销售目标和策略，提高销售效率和市场占有率；其次，需求预测有助于采购计划和安排生产计划的制定，减少受业务波动的影响。如果没有需求预测或者预测不准，公司内部很多关于销售、采购、财务预算等决策都只能根据经验而来了，会导致对市场预测不足，无法精准地制定采购、生产和投资计划，产生产能短缺，库存和资金的积压或不足等问题，增加企业库存成本。

## 1.2 问题重述

1. 根据国内某大型制造企业提供的 2015 年 9 月 1 日至 2018 年 12 月 20 日面向经销商的出货数据，分析产品价格、所在区域、节假日、促销活动和季节的对需求量的影响。同时，研究分析不同区域、不同销售方式、不同品类和不同时间段产品需求量的特性。

2. 基于问题一的分析，建立数学模型对 2019 年 1 月、2 月、3 月的月需求量分别按天、周、月的时间粒度进行预测。同时分析不同的预测粒度对预测精度会产生什么样的影响。

### 1.3 主要工作



## 2 数据预处理

### 2.1 缺失值情况说明与处理

在对数据进行分析时，缺失数据通常表现为一下三种形式，一种表现为数据的取值为 0，一种表现为数据的取值为空值，一种表现为数据的取值为特殊符号，例如取值为“-”。本次数据不存在确实情况，故不需要对数据进行缺失值处理。

### 2.2 异常值情况说明与处理

异常值是在数据集中与大部分数据明显不同的值，异常值的存在可能会对数据的分布和模型的效果产生影响，因此有必要在预测前处理数据中的异常值，避免异常值对模型的训练和预测产生不良影响。在实际数据处理过程中，我们采用箱线图对产品价格与订单需求量进行可视化，发现产品价格存在极端数据，根据  $3\sigma$  原则将产品价格的异常值删除。

### 2.3 重复值情况与处理

本文中在完成了异常值处理的基础上，对数据集进行重复值处理，数据的重复值指的是在数据集中，用于分析的各个字段均一致的数据记录。另外重复值的定义即为在数据集中用于分析的各个字段均一致。在此定义的 基础上，本文对重复值进行筛选并做删除处理。

### 3. 数据分析

在对原始数据变量进行箱线图分析时，我们发现产品价格以及订单需求量的分布存在一些离散点。这些离散点可能是由于数据收集过程中存在错误或异常情况导致的，也可能是由于数据本身存在某些特殊的特征。考虑到购买行为收到多重因素的影响，例如 2022 年“双十一”期间全网商品交易总额达到 11154 亿元，交易总额远远超于平时。这类离散点实际上反映了数据的某些特殊特征，在这种情况下，保留离散点可以帮助我们更好地理解数据的分布情况，得到更准确的分析结果。因此在解决第一个问题时，暂不对异常值做处理。

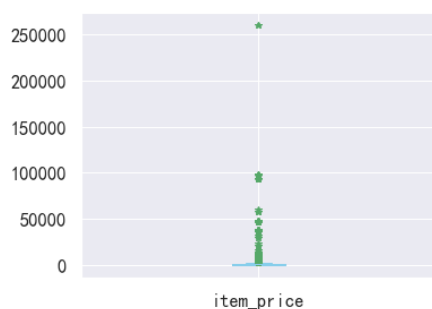


图 3-1 产品价格箱线图

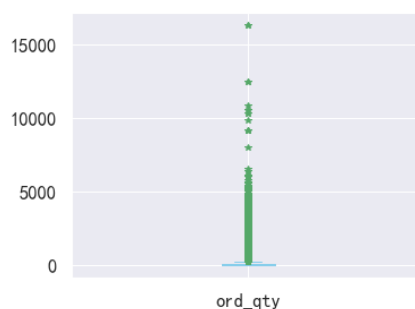


图 3-2 订单需求量箱线图

#### 3.1 价格对需求量的影响

图 3-3 和图 3-4 绘制了产品价格与需求量的散点图，并计算了两个变量的线性相关系数。通过对散点图的观察可以得到，订单需求量随着产品价格的增高而减少，线性相关系数为负，与散点图所示的结果一致。当产品价格上升时，消费者的购买力会受到影响，因此需求会减少，消费者会寻找更便宜的替代品或者降低消费水平，这正是价格弹性的体现。此外，我们还分销售区域与产品编码绘制了产品价格与需求量的散点图，固定销售区域与固定产品编码情况下绘制的散点图与前面的分析结果一致。

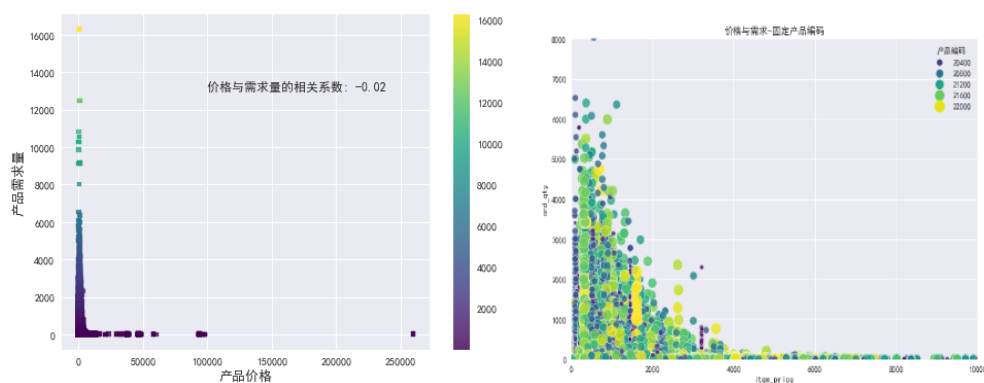


图 3-3 价格-需求量散点图

图 3-4 价格与需求量-固定产品散点图

### 3.2 区域对需求量的影响

根据图 3-5 和图 3-6 可知，不同地区的产品总需求量各不相同，其中 105 地区的需求量最高，占总需求量的 26.5%，其次是 102 地区和 101 地区，分别占总需求量的 25.5%和 22.6%。203 地区的需求量占总需求量的 21.0%，而 104 地区的需求量最低，仅占总需求量的 4.4%。这些数据反映了不同地区的市场需求量差异较大，对于企业来说，需要根据不同地区的市场情况制定相应的营销策略，以满足消费者的需求，提高产品的市场占有率。同时，也需要注意不同地区的竞争状况和消费者的消费习惯，以便更好地定位自己的产品和服务，提高市场竞争力。

对固定产品销售区域的价格需求量散点图进行观察，可以看出代表 105 号地区（黄色点）的订单数量显著多余其他地区的订单量，且该地区订单价格集中在 500 元以下及 1000-1800 元。代表 102 号地区（深蓝色点）订单的价格分布比较分散，代表 101 号地区（紫色点）订单的价格集中在 1000 元以内。在消费能力较弱的地区，消费者更加注重价格因素，在购买产品时更加注重价格的实惠性；而在消费能力较强的地区，消费者则更加注重产品的品质和服务的质量，愿意为高品质的产品和服务支付更高的价格。针对不同地区消费者的消费能力和消费需求，企业可以采取不同的销售策略。对于消费能力较弱的地区，企业可以考虑推出价格实惠、性价比高的产品，以满足消费者的需求，同时也可以采取促销和打折等方式吸引消费者的注意力；对于消费能力较强的地区，企业则可以注重产品的品质和服务的质量，提供高品质的产品和服务，以吸引高端消费者的关注，同时还可以采取差异化的定价策略，根据产品的特点和市场需求，制定不同的价格策略，以适应市场的变化和需求的变化。

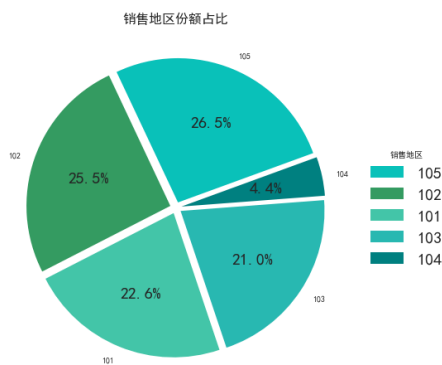


图 3-5 区域需求量饼状图

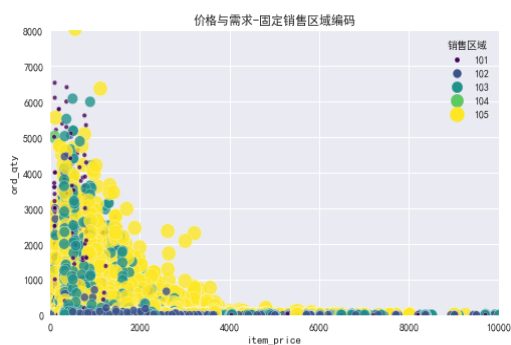


图 3-6 价格与需求量-固定销售区域散点图

### 3.3 销售方式对需求量的影响

通过图 3-7 可以看出线上销售量远远高于线下销售量，这表明消费者越来越倾向于通过线上渠



道进行购物，线上销售渠道的市场占有率相对较高。这可能是由于线上销售具有便捷、多样、价格优势等优点，吸引了更多的消费者，同时电商平台的营销推广和数据分析等技术也能够更好地满足消费者需求，提高销售效率和销售额。

接着分区域对销售渠道进行分析，通过图 3-8 可以得出不同地区的消费者在线上 and 线下购物渠道的选择存在明显的差异。其中在 102 号地区，线上销售的商品需求量线下销售的商品需求量的 40 倍以上，这类地区的消费者更倾向于在线上购物，而 104 号地区线下需求量是线上需求量的近 17 倍，这类地区则更倾向于在线下购物。影响消费者消费行为的因素很多，在一些人口密集、经济发展较好的地区，消费者更容易选择线上购物，而在一些人口分散、经济相对较弱的地区，消费者则更倾向于线下购物。此外，不同地区的消费文化也可能对购物渠道的选择产生影响，例如一些地区的消费者更注重实体店面的购物体验，而另一些地区的消费者更注重在线购物的便利性。

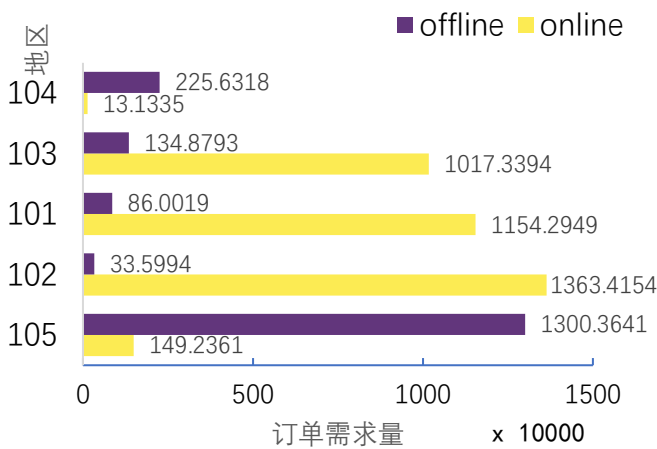


图 3-7 销售渠道与需求量-固定销售区域散点图

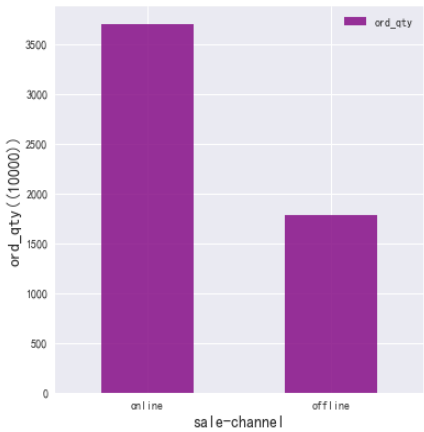


图 3-8 销售渠道-需求量条形图

此外，我们分销售渠道绘制了产品价格与需求量的散点图，从图 3-9 中可以看出，线下销售方式下产品价格与需求量之间的关系比在线上销售方式下更加密切。也就是说，当产品在线下销售时，价格的变化对需求量的影响更为显著，需求量随着价格的升高而减少的趋势更为明显。管线上购物的方便性和快捷性受到越来越多消费者的青睐，但是在购买高价位商品时，许多消费者仍然更倾向于选择线下购买。这种现象的主要原因可能是消费者在购买高价位商品时更加关注商品的品质、性能、材料等方面，需要亲身体验和了解商品的真实情况，而线下购买可以提供更加直观、全面的体验和了解，更容易让消费者做出决策。此外，在购买高价位商品时，消费者通常会考虑到商品的保修、维护等售后服务，而线下购买可以提供更加可靠、及时的售后服务，从而让消费者更加放心。

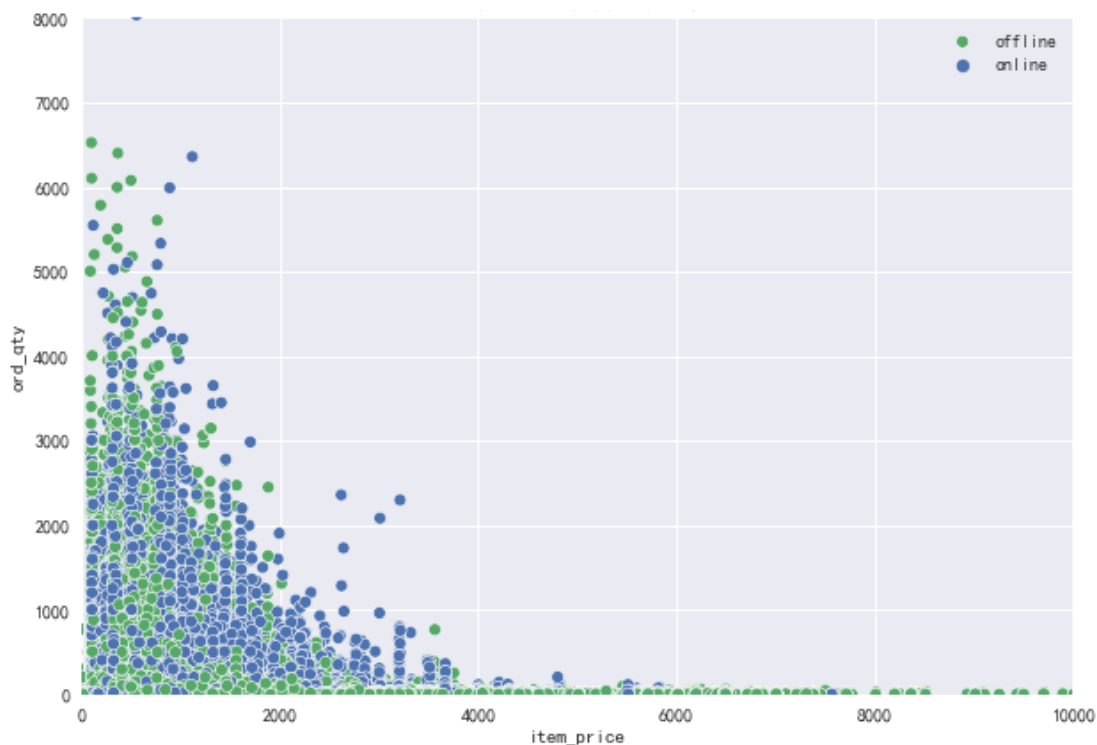


图 3-9 价格与需求量-固定销售渠道散点图

### 3.4 品类对需求量的影响

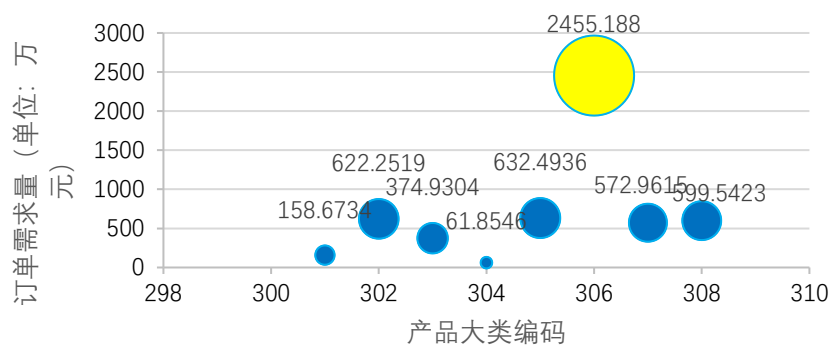


图 3-10 品类-需求量气泡图

通过图 3-10 可以得到，不同品类的需求量是存在显著下差异的，306 号产品的需求量远远高于其他品类，其需求量是 204 号产品需求量的 40 倍左右消费者。302 号、303 号、305 号 307 号与 308 号产品的需求量居中，301 号与 304 号对应的产品需求量很低。不同的产品品类通常具有不同的属性和特征，而这些属性和特征会影响消费者的购买行为和需求量。例如，对于食品品类，不同的食品可能具有不同的口味、营养成分、包装等属性，这些属性可能会影响消费者的偏好和购买行为，从而导致不同的产品品类之间的需求量差异。同样的，对于电子产品品类，不同的电子产品可能具有不同的功能、性能、品牌等属性，这些属性也会对消费者的购买行为和需求量产生影响。因此要

根据产品品类的特点和市场需求，制定不同的营销和销售策略，以提高产品的销售量和市场占有率。

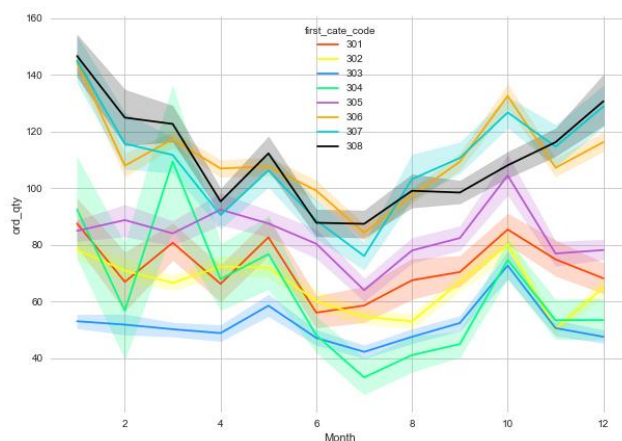


图 3-11 品类与需求量-固定月度折线图

进一步分析，尽管不同类别产品需求量存在差异，但是不同类别的产品在各月份需求的变化上基本一致，下面我们将分析时间段对产品需求的影响。

### 3.5 时间段对需求量的影响

为了探索不同品类商品的日需求变化量，我们绘制了 2015-2018 年的日需求量变化图。根据折线图，我们可以看出品类 302、303、305 和 306 的产品的日需求量变化波动较大，而其他品类的产品需求量波动范围较小。除此之外，我们还可以看出 306 品类下的产品日需求量明显高于其他产品。这说明 306 类别的产品较其他类别产品更加畅销，使得经销商订购量也更多。据此我们可以看出对于类别编码为 302、303、305 和 306 的商品，由于其波动程度大且不规则，我们的预测难度会更大。

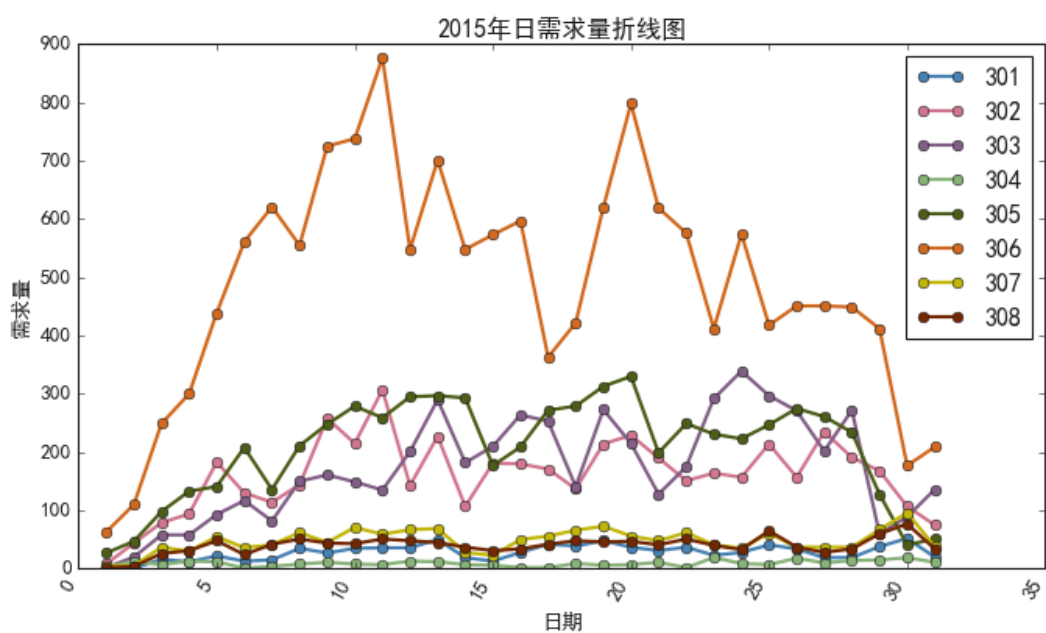


图 3-12 2015 年日需求量折线图

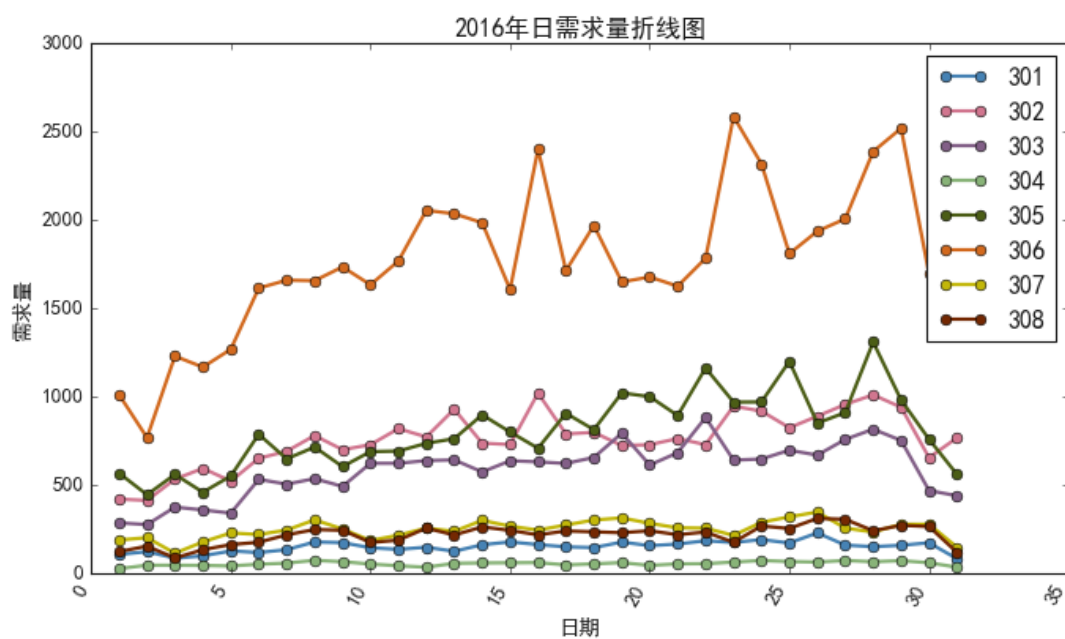


图 3-13 2016 年日需求量折线图

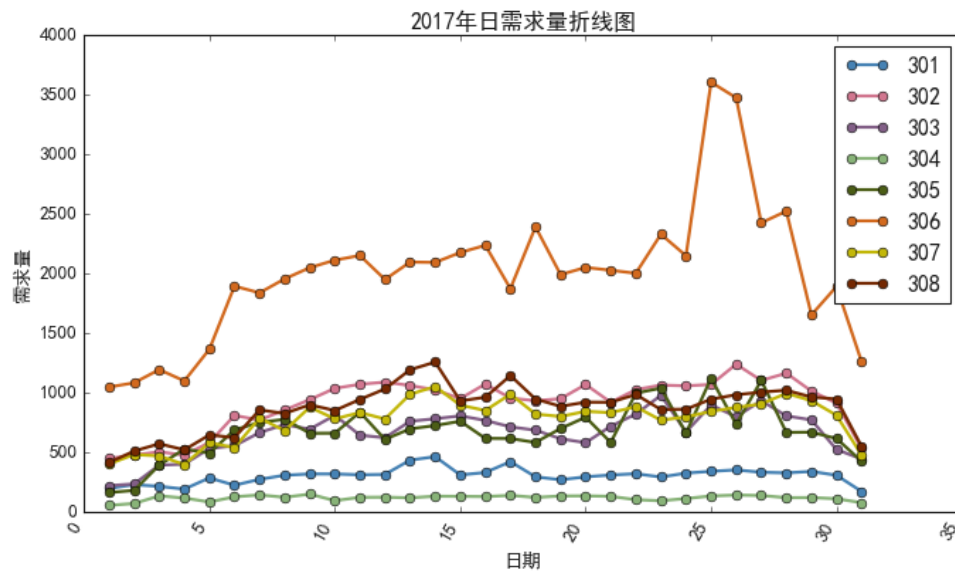


图 3-14 2017 年日需求量折线图

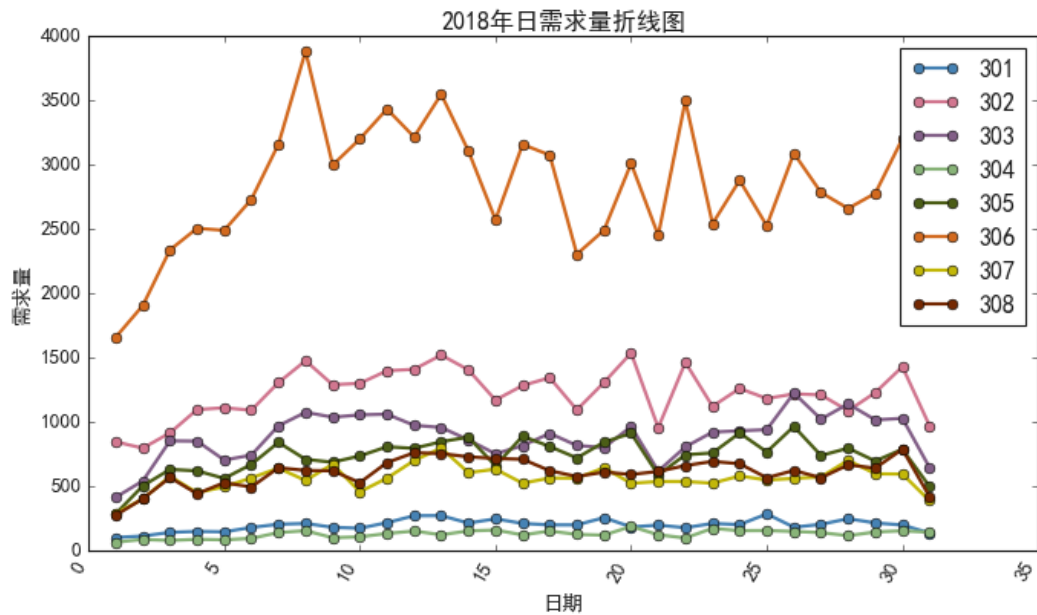


图 3-15 2018 年日需求量折线图

此外，我们绘制了 2015-2018 年商品日需求量的整体变化图。根据结果，我们可以看出 1 日至 10 日，即月初，需求量会呈现出逐步增加的趋势；在 12 日至 17 日左右，需求量处于最高的阶段；月末，即 29 日至 30 日左右，需求量呈现明显下降趋势。这可以为我们提供一个构建特征的思路，因为月中是需求量最高的时间段，所以我们可以构建一个特征来判断交易时间是否处于月中。

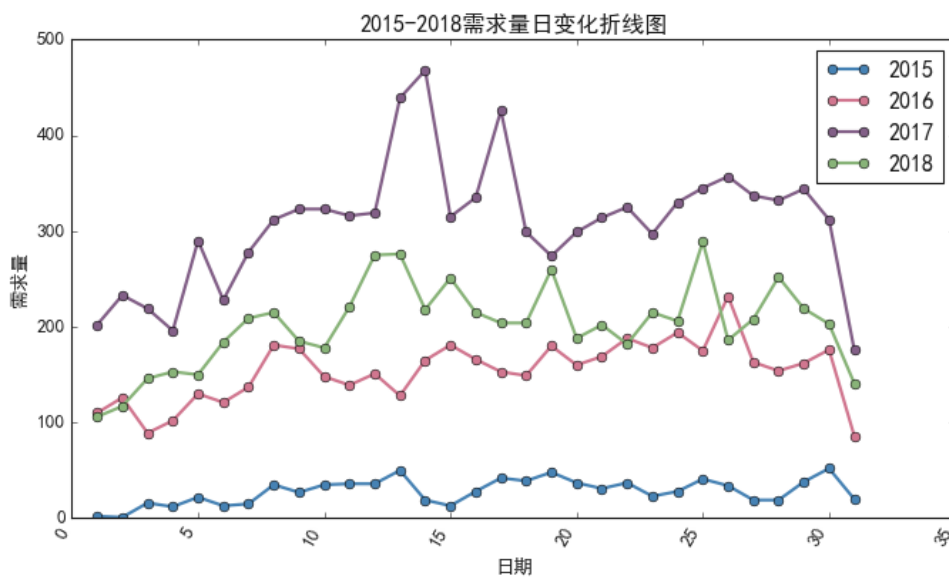


图 3-16 2015-2018 年日需求量折线图

以上是对交易日期按日期数进行整体的分析结果，接下来对需求量在各月份的特点进行分析。折线图的结果显示一年内的需求量会存在规律性波动。2016 年与 2017 年的需求量在十月份达到最高峰，而在六七月份则会出现低谷，在一月出现回升。2018 年也呈现出类似的趋势，在六七月进入低谷，十月出现大幅度回升，一月达到高峰。其中，十月或一月份达到最高峰的原因可能与节假日购物季节、年初预算安、促销活动或年终奖金有关。在这段时间里，消费者可能有更多的可支配收入，更愿意消费。而六七月份的销售低谷则可能与暑期消费习惯改变或企业活动减少有关。三月五月小范围的波动可能与庆祝活动和节日有关（如三八女神节、五一劳动节、情人节、520 等特殊日子）。企业可以充分利用历史数据，调合理安排生产计划，细化资金预算等。如在销售高峰期增加生产能力，提高产品的生产效率和质量。在销售低谷期间，适当降低生产能力，以避免库存积压和资金浪费。在销售高峰期增加促销和广告投入，提高品牌知名度和销售额。在销售低谷期间，降低广告投入，节约资金，以保证企业的正常运营和盈利等。

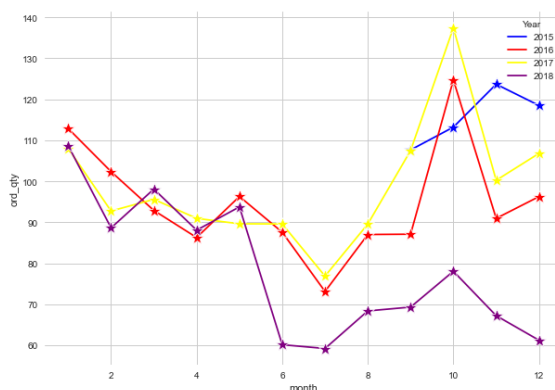


图 3-17 月份-需求量折线图

根据 3-18，可以发现大多数销售发生在周日，其次是周六和周二，而周日的销售量相对较少。

这种销售行为的规律性波动可能与消费者的购物时间和行为习惯有关。周日通常是休息日，消费者有更多的时间和机会进行购物，因此销售量相对较高。周六和周二也可能是因为消费者在周末或周初购物的需求较为集中，因而销售量也相对较高。而周一则可能因为消费者刚刚度过周末，购物需求较少，因而销售量相对较低。企业可以选择在周末投放广告，周末是消费者购物和娱乐的高峰期。不论是线上投放还是线下投放都可以吸引更多的目标受众，人们通常有更多的时间和机会进行购物和娱乐活动，广告的曝光率和转化率相对较高。

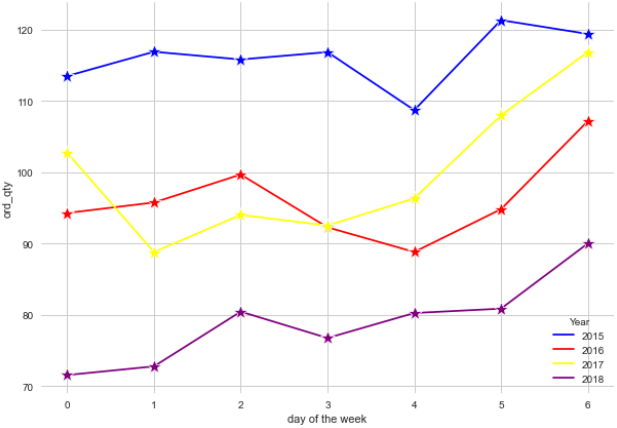


图 3-18 星期-需求量折线图

### 3.6 节假日对需求量的影响（假日：1，非假日：0）

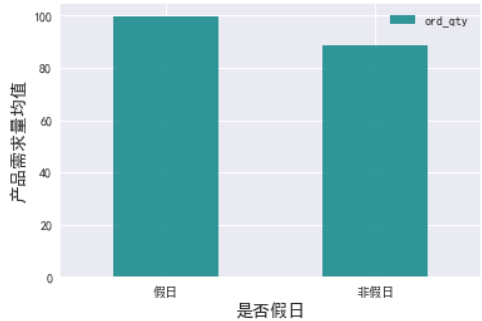


图 3-19 假日-需求量条形图

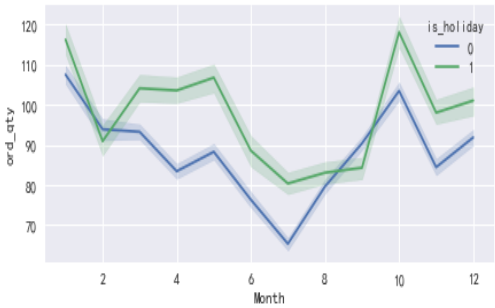


图 3-20 假日-需求量对比折线图

根据图 3-19 可以得知，节假日对需求量的影响是显而易见的，假日时人们对产品的需求高于非假日，在假日期间，人们有更多的时间和机会进行购物消费，并且在节假日，人们通常会购买一些与节日相关的商品，例如礼品、装饰品、食品等，以庆祝和纪念节日，从而带动了相关行业的销售。企业可以根据不同的节日和应节商品的特点，制定相应的促销策略和营销计划，以提高销售量和市场占有率。同时，竞争对手也会加强促销和营销力度，企业需要关注竞争对手的活动和策略，及时进行反应和调整，保持竞争优势。



### 3.7 促销对需求量的影响

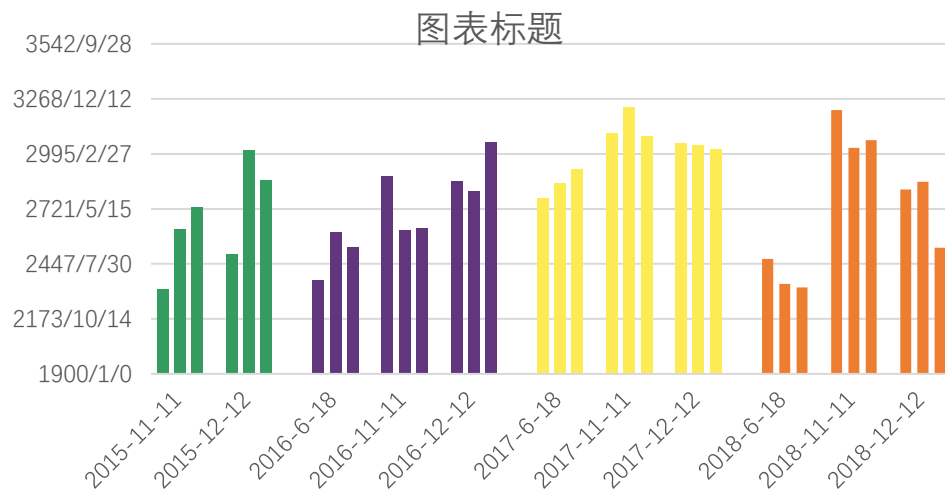


图 3-21 促销-需求量条形图

上图汇总了年中促销、双十一与双十二所在周的需求量已经其前一周与后一周的需求量。条形图显示 2015 年双十二、2016 年年中大促、2017 年双十一、2018 年双十二当周的需求量高于其前一周与后一周的需求量。很多商家大型购物节前提前展开预售活动，吸引消费者进行购物。因此，在购物节到来的前几天，消费者的购物欲望和需求已经被激发，因此销售量也会明显增加。这也是商家提前展开预售活动的原因之一，既可以提前获取销售收入，又可以为购物节的销售打下良好的基础，同时还能吸引更多消费者的关注和参与。此外，企业可以在全面高度重视的购物节期间设计一些特别的活动和优惠，如限时抢购、满减、秒杀等，增加销售量和品牌知名度；提前准备充足的库存，以保证商品供应的稳定性和销售量的增长，值得一提的是，随着电商的发展，企业需提前规划物流配送方案，确保商品能够及时、准确地送达消费者手中，并提供优质的售后服务。

### 3.8 季节对需求量的影响

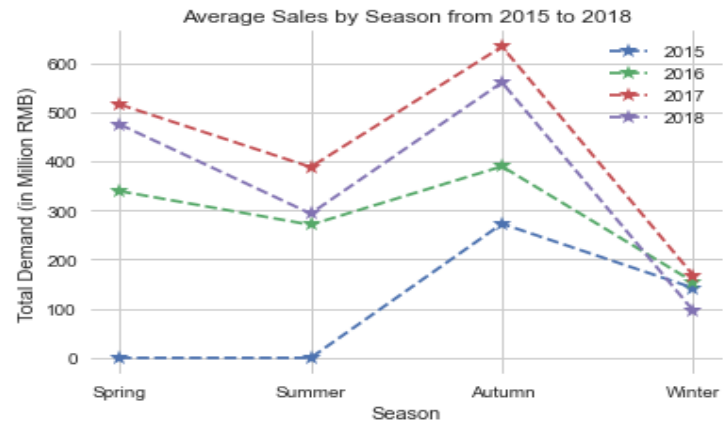




图 3-22 季节-需求量折线图

根据折线图可以看出，订单需求量在不同季节之间呈现出一定的差异。具体而言，秋季的订单需求量最高，其次是春季，而冬季的订单需求量最低。春季和秋季是人们消费需求旺盛的季节，因为在这两个季节，气温适宜，人们的活动量相对较大，同时也有一些重要的节日和购物热点，例如中秋节、国庆节、春节等，这些都会带动相关产品的销售，自然也会有更多的消费需求和购买行为。而在冬季，由于气温较低，人们的活动量减少，消费需求也相应下降，因此订单需求量也相对较低。这种规律性波动表明存在季节性因素或市场趋势影响产品或服务的需求量。企业可以把握需求量的规律性波动，预测和应对市场变化，提高销售效益和企业竞争力。如制定针对不同季节性波动的销售计划和策略，包括促销活动、营销推广、库存管理等。这些计划和策略可以根据历史数据和市场趋势进行调整和优化。

## 4 特征工程构建

特征工程的作用主要是依据人类经验从原始数据中提取信息，并转化成计算机可以理解语言传递给模型，其本质为人类经验的数字化<sup>[1]</sup>。

### 4.1 特征构造

时间序列数据一般情况下可以划分为以下几类：时间戳、时序值以及时间序列的属性变量。本题中的订单日期即为时间戳，需求量即时序值，而销售区域编码、产品编码、产品类别及产品价格即为序列的属性变量。本文特征工程也围绕这三类数据产生。

#### 4.1.1 时间戳衍生特征

时间戳数据可以衍生出以下几类特征。

(1) 时间特征

我们可以直接从日期中提取时间特征，比如交易发生的时间、季度、月份和日期。本节所构造的特征如表 4-1 所示。

表 4-1 特征及特征解释

时间特征	year	交易发生年份
	month	交易发生月份
	day	交易发生的天数（对应于月）
	quarter	交易发生的季度
	dayofyear	交易发生的天数（对应于年）
	week	交易发生日期属于该年度的周数

(2) 布尔特征

结合真实场景，订单交易发生的时间与是否处于工作日或节假日有关。由此，我们可以构造以下特征：

1. 是否处于工作日

企业下单时间一般处于工作日，而在周末休息时订单量一般会大量减小，因此构造特征

is\_work\_day 来判断是否处于可能发生交易的时期。根据提取后的特征绘制饼状图，如图 4-1 所示，可以看出工作日需求量占比明显大于非工作日占比。

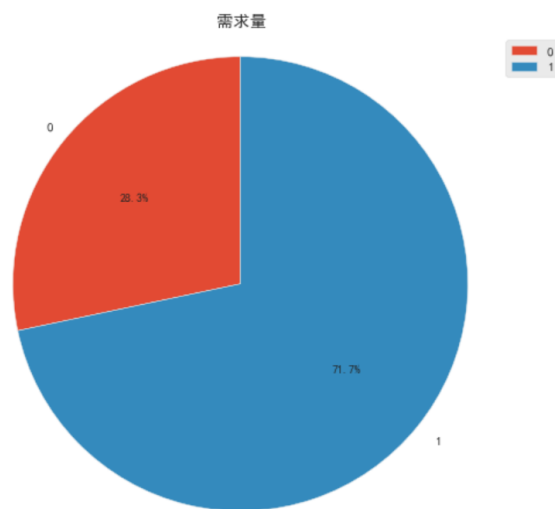


图 4-1 是否是工作日对需求量的影响

## 2. 是否处于节假日

一般情况下，企业在国家法定节假日期间不会产生订单，这就说明需求量可能在节假日期间出现明显减少的现象。通过查询国务院办公厅发布的通知，以下为 2015 年-2019 年的部分节假日安排。

### 2015 年：

- 一、元旦：1月1日至3日放假调休，共3天。1月4日（星期日）上班。
- 二、春节：2月18日至24日放假调休，共7天。2月15日（星期日）、2月28日（星期六）上班。
- 三、清明节：4月5日放假，4月6日（星期一）补休。
- 四、劳动节：5月1日放假，与周末连休。
- 五、端午节：6月20日放假，6月22日（星期一）补休。
- 六、中秋节：9月27日放假。
- 七、国庆节：10月1日至7日放假调休，共7天。10月10日（星期六）上班。

### 2016 年：

- 一、元旦：1月1日放假，与周末连休。
- 二、春节：2月7日至13日放假调休，共7天。2月6日（星期六）、2月14日（星期日）上班。
- 三、清明节：4月4日放假，与周末连休。
- 四、劳动节：5月1日放假，5月2日（星期一）补休。
- 五、端午节：6月9日至11日放假调休，共3天。6月12日（星期日）上班。
- 六、中秋节：9月15日至17日放假调休，共3天。9月18日（星期日）上班。
- 七、国庆节：10月1日至7日放假调休，共7天。10月8日（星期六）、10月9日（星期日）上班。

### 2017 年：

一、元旦：1月1日放假，1月2日（星期一）补休。

二、春节：1月27日至2月2日放假调休，共7天。1月22日（星期日）、2月4日（星期六）上班。

三、清明节：4月2日至4日放假调休，共3天。4月1日（星期六）上班。

四、劳动节：5月1日放假，与周末连休。

五、端午节：5月28日至30日放假调休，共3天。5月27日（星期六）上班。

六、中秋节、国庆节：10月1日至8日放假调休，共8天。9月30日（星期六）上班。

2018 年：

一、元旦：1月1日放假，与周末连休。

二、春节：2月15日至21日放假调休，共7天。2月11日（星期日）、2月24日（星期六）上班。

三、清明节：4月5日至7日放假调休，共3天。4月8日（星期日）上班。

四、劳动节：4月29日至5月1日放假调休，共3天。4月28日（星期六）上班。

五、端午节：6月18日放假，与周末连休。

六、中秋节：9月24日放假，与周末连休。

七、国庆节：10月1日至7日放假调休，共7天。9月29日（星期六）、9月30日（星期日）上班。

2019 年：

一、元旦：2018年12月30日至2019年1月1日放假调休，共3天。2018年12月29日（星期六）上班。

二、春节：2月4日至10日放假调休，共7天。2月2日（星期六）、2月3日（星期日）上班。

三、清明节：4月5日放假，与周末连休。

四、劳动节：5月1日放假。

五、端午节：6月7日放假，与周末连休。

六、中秋节：9月13日放假，与周末连休。

七、国庆节：10月1日至7日放假调休，共7天。9月29日（星期日）、10月12日（星期六）上班。

根据以上信息构造 `is_special_day` 特征，判断交易日期是否处于节假日。绘制提取后的特征的需求量饼状图，如图 4-2 所示，可以看到节假日期间需求量占比极小。

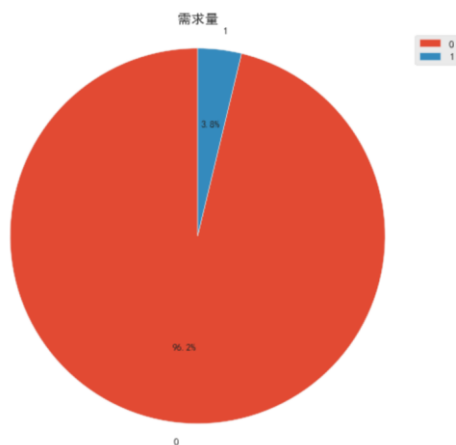


图 4-2 是否是节假日对需求量的影响

### 3. 是否处于月初或月末

对于各经销商来说，月初和月末是对一个月的工作结果进行归集、编制报表和纳税汇报的时期。所以月初和月末是经销商较为忙乱的一个时期，据此可以推测各经销商更有可能下单的时期位于月中，即月中的订单需求量会更大。提取 `is_month_start` 和 `is_month_end` 特征来判断交易时间是否位于月初和月末。将提取出的特征绘制需求量饼状图如图 4-3 所示。如图所示，月初和月末的需求量占比极低，结果符合我们的推测。即月中订单需求量更高。

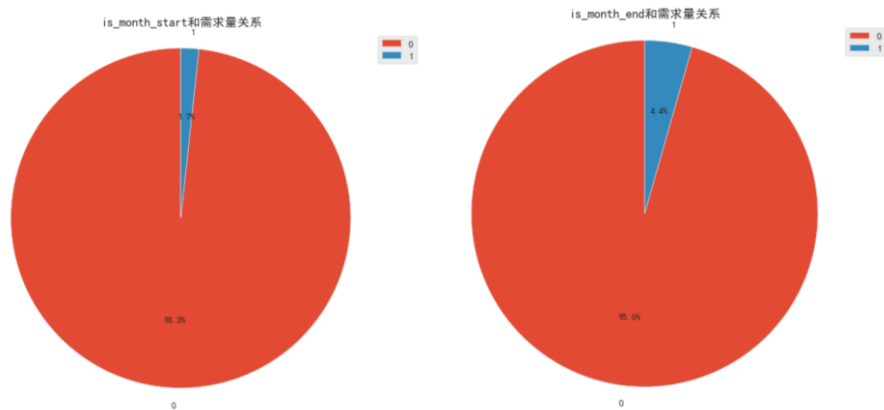


图 4-3 月初或月末对需求量影响

本节构造的特征如表 4-2 所示。

表 4-2 特征及特征解释

布尔特征	is_sales	交易日期是否处于发生促销的时期
	is_work_day	交易日期是否处于在工作日
	is_special_day	交易日期是否处于在法定节假日
	is_month_start	交易日期是否处于月初
	is_month_end	交易日期是否处于月末
	is_special_week	交易日期所在星期是否存在法定节假日

#### 4.1.2 时序值衍生特征

时间序列预测是通过历史数据对未来进行预测，那么对于当前时间点来说，过去所发生的历史数据所提供的信息十分重要。本题中的时序值为需求量。本文通过时间滑动窗口统计进行时序值衍生特征提取。滑动窗口统计即基于某段时间窗统计数据情况，比如我们选择滑动窗口为 3 天，那么

可以衍生出的特征变量即为过去 3 天内需求量的最大值、最小值、均值等统计特征。本文将滑窗设置为 28，即统计过去 28 天内需求量的相关信息。对于日时间粒度预测，本文将滑窗大小设置为 30；对于周时间粒度预测，本文将滑窗大小设置为 4；对于月时间粒度预测，本文将滑窗大小设置为 1。

本节构造的特征如表 4-3 所示。

表 4-3 特征及特征解释

时序值特征	lag_1	1 个月前的历史需求量数据
	rmean_1_1	1 个月前的历史需求量数据的移动平均值
	lag_4	4 周前的历史需求量数据
	rmean_4_4	4 周前的历史需求量数据的移动平均值
	lag_30	30 天前的历史需求量数据
	rmean_30_30	30 天前的历史需求量数据的移动平均值

### 4.1.3 属性变量

本文的属性变量主要是产品大类编码、产品细类编码等类别编码。由于这些类别并不反应顺序，可以将这些类别变量进行 label encoding。此外还对这个产品类别下的需求量均值、标准差进行提取作为新特征。

此外由于测试集中并不包含产品价格这一特征，但是对于产品订单需求来说，价格是一个相对重要的特征。因此本文将训练集中同一类别下的同一个产品的价格进行平均作为测试集对应产品的价格特征。

此外对价格进行分箱处理。之所以要进行分箱处理，是为了实现数据的离散化，增强数据稳定性，减少过拟合的风险。经过分箱后，特征出现模型无法处理的值的可能性就降低，这样既提高了模型的泛化能力，也会加快模型的训练速度。

本节构造的特征如表 4-4 所示。

表 4-4 特征及特征解释

属性变量衍生特征	brand_qty_max	同一大类下的商品需求量最大值
	brand_qty_median	同一大类下的商品需求量中位数
	brand_qty_min	同一大类下的商品需求量最小值

	brand_qty_sum	同一大类下的商品需求量总值
	brand_qty_average	同一大类下的商品需求量平均值

## 4.2 特征选择

特征选择的目的是通过去除原始特征与构造的新特征中的冗余和无效特征，以达到数据降维的目的。有效的特征选择方法可以减少模型的学习时间，并可以在一定程度上提高模型精度<sup>[2]</sup>。本文采用 Sequential forward selection (SFS)的方法进行特征选择。

SFS 又称前向搜索法，其主要思想类似于枚举法。SFS 先从所有特征中挑选一个最优特征，记作特征 $a$ ；然后在从头开始继续搜索，寻找一个特征 $b$ ，这个特征 $b$ 和特征 $a$ 组合在一起可以达到最优的分类效果，以此类推。

经过筛选后的特征如表 4-5 所示。

表 4-5 特征及特征解释

筛选后特征	sales_region_code	销售区域编码
	first_cate_code	产品大类编码
	second_cate_code	产品细类编码
	year	交易发生的年份
	month	交易发生的月份
	week	交易发生的周数
	item_price	商品价格
	brand_qty_median	同一大类下的商品需求量中位数
	brand_qty_average	同一大类下的商品需求量均值
	rmean_1_1	1 个月前的历史需求量数据的移动平均值
	lag_4	4 周前的历史需求量数据
	rmean_4_4	4 周前的历史需求量数据的移动平均值
	lag_30	30 天前的历史需求量数据

	rmean_30_30	30 天前的历史需求量数据的移动平均值
--	-------------	---------------------



## 5. 预测算法原理与评价指标

### 5.1 BiLSTM 算法原理

BiLSTM (Bidirectional Long Short-Term Memory) 是一种双向长短时记忆网络 (LSTM)，它是一种递归神经网络，主要用于处理序列数据。与传统的 LSTM 不同的是，BiLSTM 是前向 LSTM 和后向 LSTM 的组合，用于学习双向长期依赖关系<sup>[3]</sup>。图 5-1 展示了 BiLSTM 的简要结构。

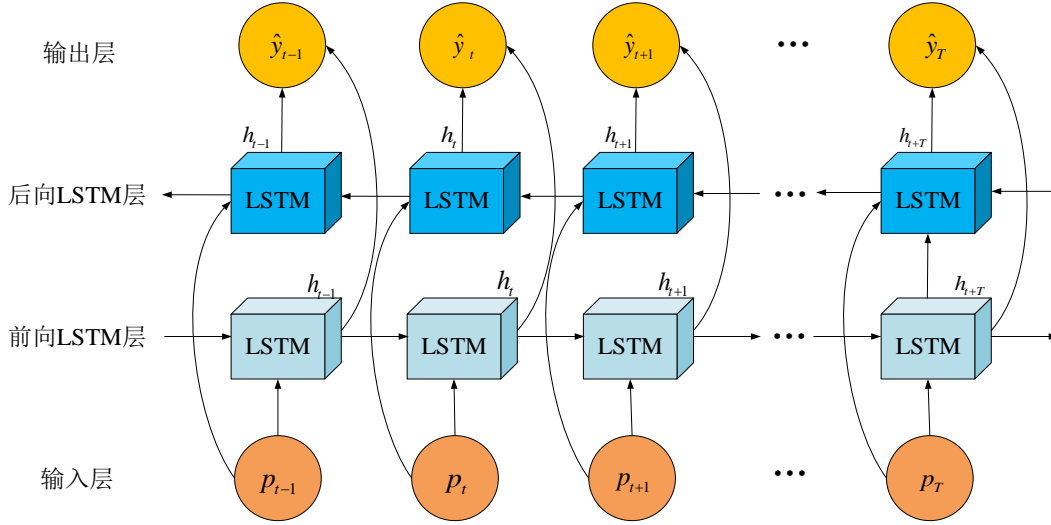


图 5-1 BiLSTM 的简要结构

在 BiLSTM 中，前向 LSTM 的输入为时间步从 1 到  $T$  的序列，而后向 LSTM 的输入为时间步从  $T$  到 1 的序列。对于  $t$  时刻的输入  $x_t$ ，前向 LSTM 和后向 LSTM 的输出  $\vec{h}_t$  和  $\overleftarrow{h}_t$  可以通过标准 LSTM 计算。LSTM 以上一时刻的细胞状态  $C_{t-1}$ ，隐藏层状态  $h_{t-1}$  和当前时刻的输入  $x_t$  作为三个门的输入。

首先，遗忘门计算决定因子  $f_t$  来确定细胞状态  $C_{t-1}$  中多少信息应该得到保留。

$$f_t = \sigma(w_f \otimes [h_{t-1}, p_t] + b_f)$$

其中， $w_f$  和  $b_f$  分别为遗忘门的权重和偏置值。 $\sigma(\cdot)$  表示 sigmoid 函数，它将输出值转化至 0-1 范围内， $f_t$  越接近 1 表明信息越应该保留，越接近 0 则越应该舍弃。

其次，输入门将细胞状态进行更新，如下所示。

$$\hat{C}_t = \tanh(w_c \otimes [h_{t-1}, p_t] + b_c)$$

$$i_t = \sigma(w_i \otimes [h_{t-1}, p_t] + b_i)$$

$$C_t = f_t \otimes C_{t-1} + i_t \otimes \hat{C}_t$$

当前细胞状态  $\hat{C}_t$  记录输入中的信息， $\tanh(\cdot)$  激活函数将输出值转化为-1 到 1 之间的数， $w_c$  和  $b_i$  为计算  $\hat{C}_t$  权重和偏置值。类似于遗忘门的计算过程，输入门计算决定因子  $i_t$  来保留  $\hat{C}_t$  中的信息， $w_i$  和  $b_i$  为计算  $i_t$  的权重和偏置值。 $\hat{C}_t$  将与  $C_{t-1}$  合并为新的细胞状态  $C_t$ 。

最后，输出门通过决定因子  $o_t$  来控制应该输出多少信息并传递给下一时刻的记忆单元。

$$o_t = \sigma(w_o \otimes [h_{t-1}, p_t] + b_o)$$

$$h_t = o_t \otimes \tanh(C_t)$$

一下三个公式可以记作标准 LSTM 函数  $LSTM(\cdot)$ ，则 BiLSTM 的输出可通过下式计算。

$$\overset{\text{f}}{h}_t = LSTM(\overset{\text{f}}{h}_{t-1}, p_t)$$

$$\overset{\text{s}}{h}_t = LSTM(\overset{\text{s}}{h}_{t-1}, p_t)$$

$$\hat{y}_t = \varphi_B(\overset{\text{f}}{w}_t \otimes \overset{\text{f}}{h}_t + \overset{\text{s}}{w}_t \otimes \overset{\text{s}}{h}_t + b_t)$$

其中， $\hat{y}_t$  表示 BiLSTM 的预测值， $\overset{\text{f}}{w}_t$  和  $\overset{\text{s}}{w}_t$  分别为前向隐藏层和后向隐藏层的权重值， $b_t$  为 BiLSTM 输出层的偏置值。

## 5.2 BiLSTM-Attention 算法原理

BiLSTM-Attention 是一种能够应用于序列到序列 (Seq2Seq) 任务的神经网络结构，其主要目的是提取序列中的重要特征并将其加以利用。BiLSTM-Attention 模型由两个主要部分组成：双向长短期记忆网络 (BiLSTM) 和注意力机制 (Attention)。BiLSTM 能够有效地捕捉序列中的上下文信息，而 Attention 机制则能够根据不同的输入情况动态地调整模型关注的部分，这能使模型在长时间保留信息的基础上，自动发现那些对于产品订单需求量起到关键作用的特征，根据 Attention 对象的重要程度，重新分配权重，突出某些重要特征，从而有侧重地对历史数据加以利用，避免了传统的任务中复杂的特征工程，使模型更加易于训练和部署。

BiLSTM-Attention 的模型结构如图 5-2<sup>[4]</sup>所示：

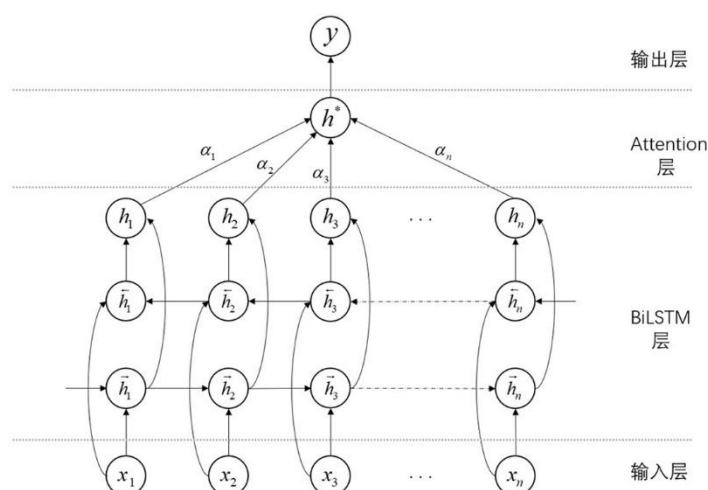


图 5-2. BiLSTM-Attention 模型结构

### 5.3 LightGBM 算法原理

LightGBM (Light Gradient Boosting Machine) 是一种基于决策树的梯度提升框架，由微软公司开发。相比于传统的 GBDT (Gradient Boosting Decision Tree) 算法，LightGBM 具有更高的训练效率、更低的内存消耗、更好的准确率和可扩展性等优点。其核心思想是基于梯度提升决策树 (GBDT)，将  $M$  个弱的回归树依次合并成一个强大的回归树<sup>[5]</sup>。其计算公式如下：

$$F(x) = \sum_{m=1}^M f_m(x)$$

垂直生长和直方图算法是 LightGBM 模型的主要创新点。传统的决策树算法是水平生长的，即一次性找到最优的分裂特征和分裂点。而 LightGBM 采用垂直生长，即先生成所有叶子节点，再根据一个评价指标（如信息增益或基尼系数）来选择最佳的分裂特征和分裂点，从而减少了计算量。此外，传统的决策树算法需要对每个特征的每个取值都进行搜索，计算其增益，这样非常耗时。而 LightGBM 采用直方图算法来构建决策树，将连续的特征离散化成若干个离散的特征值，然后对这些特征值建立直方图，计算直方图上的增益来选择最佳的分裂点。这样可以大大减少计算量，提高了训练速度。

LightGBM 使用直方图法合并了相互独特的特征。其基本概念是在创建直方图之前将连续特征值分为  $M$  个积分。根据直方图的离散值系统，对数据进行导航，找到决策树。因为决策树是一个弱模型，直方图算法大大减少了计算时间，这种分割方法常会取得更好的结果。

另外在 LightGBM 中，采用的是 Leaf-wise 方法，从所有现有的叶子中找出散射增益最高的叶子，然后将其分割。Leaf-wise 的缺点是，它可能通过生长一棵彻底的决策树而产生过拟合。LightGBM

用最大深度限制来增强 Leaf-wise，以确保最大效率，同时试图避免过度拟合。

## 5.4 集成学习与组合算法原理

集成学习（Ensemble Learning）是一种机器学习技术，是将多个弱学习器（又称个体学习器）训练结果采取一定的结合策略融合成一个强学习器的过程，组合算法（Combination Algorithm）是集成学习中用于组合多个弱学习器的算法，即结合策略。集成学习有两个主要步骤：首先是确定个体分类器的建立方式，然后采用特定策略结合各个个体分类器，集成学习的具体流程如图 5-3 所示。它使用多种学习算法组合来获得比单独使用任何一种学习算法都要好的预测性能。

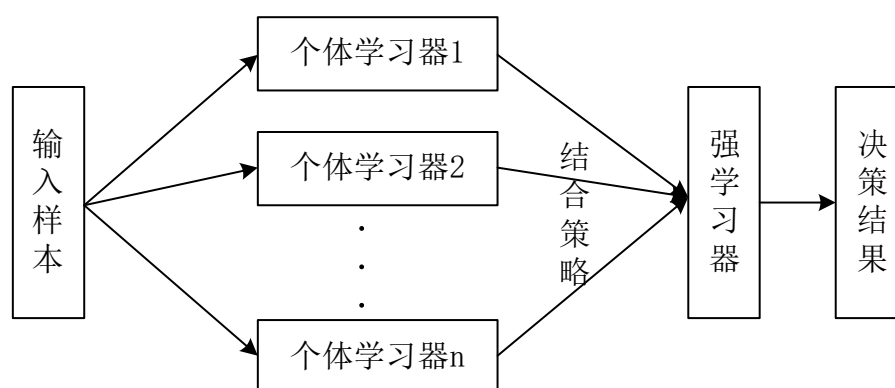


图 5-3 集成学习算法流程图

个体学习器通常由一个现有的学习算法从训练数据产生，例如支持向量机算法、C4.5 决策树算法、BP 神经网络算法等，此时若个体学习器为同一类别，这样的集成为同质集成，若个体学习器为不同类型，成为异质集成。另一方面，根据个体学习器的组合模式划分为串行式和并行式两种方式，串行式集成学习中，个体学习器是按照一定的顺序依次构建的，每个个体学习器的输出都会对下一个个体学习器的学习产生影响。最典型的串行式集成学习算法是 Boosting 算法，比如 AdaBoost 和 Gradient Boosting 算法。并行式集成学习中，个体学习器是同时构建的，每个个学习器之间是独立的，不会相互影响。最典型的并行式集成学习算法是 Bagging 算法和随机森林算法<sup>[6]</sup>。集成学习方法通常比单一分类器具有更为优异的泛化能力，可有效避免过拟合问题，通过集成策略可以平均不同的分类结果，从而减少选择错误预测的风险，降低模型的出错率，提高整体的预测性能。

本文使用的组合算法采用了并联组织关系，A 算法和 B 算法分别独立地预测完商品订单量后，给这两个算法赋予一定的加权系数，相加计算得到组合的预测结果。假设需将 I 个算法组合，第 i 个算法的权重为  $w_i$ ，预测结果为  $y_i$ ，其计算公式如下：

$$y = \sum_{i=1}^I w_i \cdot y_i$$

对于加权系数的计算方法，本文参考了“预测均方根误差倒数法”，基本思路为：若一个单一预测模型的预测误差较大，表明该模型的预测精度较低，则降低其在组合预测中的加权系数<sup>[7]</sup>。

其计算公式为：

$$w_i = \frac{e_i^{-1}}{\sum_{i=1}^m e_i^{-1}}, i = 1, 2, \dots, m$$

$$\sum_{i=1}^m w_i = 1, w_i \geq 0, i = 1, 2, \dots, m$$

其中， $e_i$ 是第  $i$  个单一预测模型的预测均方根误差。

通过“预测误差倒数法”来计算参考加权系数，设置若干组与参考加权系数相近的加权系数组合实验，比较组合算法的预测误差，得到较优的加权系数。

## 5.5 预测算法误差评价指标

由于商品需求量的预测受到各种因素的影响，我们无法做到面面俱到，会不可避免地产生一定的误差。但我们需要尽量降低误差，提高精度。本文所用到的模型多为回归预测模型，对此类模型进行客观评价的常用的误差评价指标有以下几种：

(1) 均方误差 (MSE)：观测值与真值偏差的平方和与观测次数比值的平方根，用于衡量观测值同真值之间的偏差。指标越小，模型精度越高。计算公式如下：

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

(2) 均方根误差 (RMSE)：是均方误差的算术平方根，开方后的误差结果与数据是一个级别的，它对测量中的离群点反映非常敏感。计算公式如下：

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}$$

(3) 平均绝对误差 (MAE)：是预测值与真实值的绝对误差的平均数，通过将所有样本预测的误差绝对值除以样本数量得到。此指标越小，模型的预测精度越高。计算公式如下：

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|$$

(4) 平均绝对百分比误差 (MAPE)：是样本的平均绝对误差除以样本真实值，对平均绝对误差做了归一化，为样本间的误差提供了统一衡量标准。此指标对相对误差敏感，目标变量的全局缩放不会改变它的值，适合目标变量量纲差距较大的情况。计算公式如下：

$$\text{MAPE} = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right|$$

## 6 产品需求量预测

产品需求量的预测是一个回归任务，本文分别按月、周、日时间粒度来预测。用 2018 年 9 月之前的数据作为训练集，考虑到数据集记录的最后一天是 2018 年 12 月 20 日，不满一整个月，因此用 2018 年 9 月至 2018 年 11 月三月的数据作为验证集。

每种粒度的预测都使用多个机器学习和深度学习的模型，比较它们在验证集中的预测精度即预测误差，找出在该粒度下表现最好的算法模型，并记录这个模型的预测精度。再比较三种粒度下的预测精度，选用预测精度最高的颗粒度和模型来预测未来三月的产品需求量。

### 6.1 按月时间粒度预测

#### 6.1.1 数据集处理

通过特征工程整理了名为 `month_sum.csv` 的数据集，按时间顺序排序，数据集集中的每个样本代表某商品在某地区某个月的需求情况，用 2018 年 9 月之前的数据作为训练集，2018 年 9 月至 2018 年 11 月三月的数据作为验证集。

#### 6.1.2 单/多变量 BiLSTM 模型构建与比较

本次实验用 Keras 框架实现对 BiLSTM 模型的搭建。Keras 对较复杂的 TensorFlow API 进行了封装，框架整体较为简洁便于深度网络学习模型的搭建。在数据的处理上，调用 sklearn 库的 `MinMaxScaler` 方法在模型运行前对数据做标准化处理，来提升模型收敛速度和预测精度。在计算预测误差时，再将数据反归一化计算预测误差。

由于设备的限制，首先不在本节进行过多的超参数实验，依据调研和经验设置各类超参数，比较出单变量 BiLSTM 和多变量 BiLSTM 的预测精度，选择一种模型在下一节进行调参。

通过调研，通常二层 BiLSTM 即可取得较好的预测效果<sup>[8]</sup>。激活函数选用了双曲正切 `tanh` 函数，可以将输入值转换为 -1 至 1 之间，并使输出和输入能够保持非线性单调上升和下降关系，在一定程度上减轻了梯度消失的问题，容错性好。

单变量和多变量 BiLSTM 模型的区别在于，单变量仅把需求量作为模型的输入，多变量把需求量和和其他样本特征一起输入模型训练。两模型的结构如下图 6-1，图 6-2 所示，两模型设置相同的超参数，超参数设置如表 6-1 所示：

Model: "model"

Layer (type)	Output Shape	Param #
input_1 (InputLayer)	[(None, 30, 1)]	0
bilstm (Bidirectional)	(None, 30, 200)	81600
dropout (Dropout)	(None, 30, 200)	0
bilstm2 (Bidirectional)	(None, 200)	240800
dropout_1 (Dropout)	(None, 200)	0
dense (Dense)	(None, 1)	201

=====  
 Total params: 322,601  
 Trainable params: 322,601  
 Non-trainable params: 0

图 6-1 单变量 BiLSTM 模型结构图

Model: "model\_1"

Layer (type)	Output Shape	Param #
input_2 (InputLayer)	[(None, 30, 11)]	0
bilstm (Bidirectional)	(None, 30, 200)	89600
dropout_2 (Dropout)	(None, 30, 200)	0
bilstm2 (Bidirectional)	(None, 200)	240800
dropout_3 (Dropout)	(None, 200)	0
dense_1 (Dense)	(None, 1)	201

=====  
 Total params: 330,601  
 Trainable params: 330,601  
 Non-trainable params: 0

图 6-2 多变量 BiLSTM 模型结构图

表 6-1 单/多变量 BiLSTM 模型超参数设置

超参数	超参数设置
Epoch	21
BiLSTM 激活函数	tanh
Dropout	0.01
优化器	Nadam
BiLSTM 网络层数	2
BiLSTM 网络神经元数	100
滑动时间窗口	30

该设置下，单变量和多变量 BiLSTM 模型的预测精度如表 6-2 所示，可以看出单变量与多变量 BiLSTM 模型的预测误差相差不大，但考虑到多变量模型输入的特征更多，模型可解释性更强。因此，在下一节针对多变量 BiLSTM 进行超参数调整。

表 6-2 月粒度单/多变量验证集预测误差

预测误差指标	单变量 BiLSTM	多变量 BiLSTM
mae	1213.424675	1334.905124
mape	38.21214223	47.03639145



mse	6877186.311	6847986.82
rmse	2622.439	2616.865839

### 6.1.3 多变量 BiLSTM 模型调参

多变量 BiLSTM 模型涉及很多超参数的优化，本节主要对以下超参数进行优化，以确定表现最佳的 BiLSTM 网络模型的结构和参数。为了防止神经网络训练过程中，权重的随机初始化造成调参与预测结果的随机性增大，不便于我们找到最优的参数及模型，引入了固定随机种子以准确选择参数，增加实验结果的可复现性。

在 Epoch 参数设置上，运用了早停机制 EarlyStopping 与人工调整相结合的方式。设置一个较大的 epoch，担当模型的损失值长时间不再随着迭代减小时，早停机制让训练提前结束。模型的优化器选择了 Nadam 算法，它在 Adam 算法的基础上加入了对历史梯度信息的加权，能够取得更好的优化效果。模型超参数的设置如表 6-3 所示，表格底部统计了不同超参数组合的预测误差：

表 6-3 月粒度 BiLSTM 模型超参数设置和预测误差

		设置 1	设置 2	设置 3	设置 4
超参数	Epoch	23	50	21	23
	BiLSTM 激活函数	tanh	tanh	tanh	tanh
	Dropout	0.01	0.01	0.01	0.01
	优化器	Nadam	Nadam	Nadam	Nadam
	BiLSTM 网络层数	2	2	2	2
	BiLSTM 网络神经元数	100	200	200	200
	滑动时间窗口	30	60	3	240
	batch size	128	480	256	480
验证集	mae	1242.120781	1252.534875	1253.191319	1129.760878
预测误差	mape	39.88766078	40.72918063	41.28139989	30.78954833

mse	6845330.348	6780898.247	6851515.943	6911224.429
rmse	2616.358222	2604.015792	2617.540056	2628.920773

观察各预测误差评价指标，并以 RMSE 为主要的评价指标，比较可得表 6-3 的设置 2 的超参数设置较好。图 6-3 是设置 2 的模型迭代过程的函数损失图像，可以看出模型收敛较快，算法有效。但是观察图 6-4 中模型对验证集中需求量较大的情况拟合效果不佳。

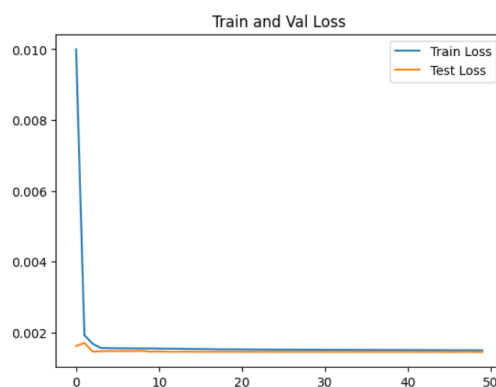


图 6-3 多变量 BiLSTM 设置 2 的损失函数图像

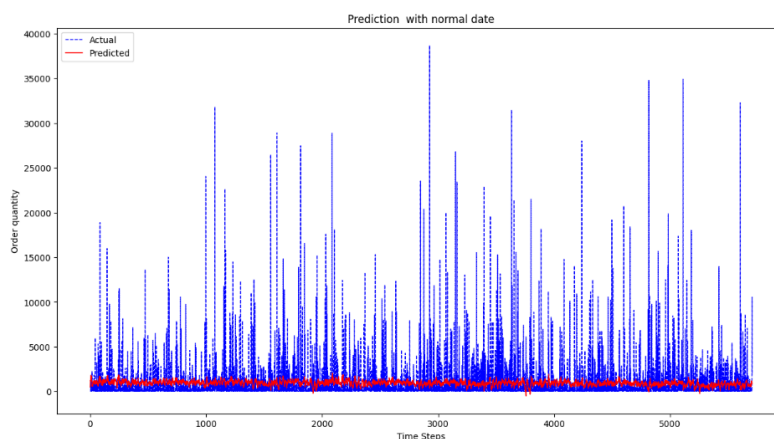


图 6-4 多变量 BiLSTM 设置 2 的验证集真实值、预测值拟合效果

#### 6.1.4 多变量 BiLSTM-Attention 模型构建

多变量 BiLSTM-Attention 模型的搭建，是在前文多变量 BiLSTM 的基础上，增加了一个激活函数为 sigmoid 的 attention 层，来优化模型的权重。此模型的各项超参数参考上文多变量 BiLSTM 的设置 2 的各项超参数设置，但在模型结构上对 BiLSTM-Attention 融合模型中 BiLSTM 层的数量进行了探索。

1 层 BiLSTM 和 1 层 Attention 的模型结构如图 6-5 所示，2 层 BiLSTM 和 1 层 Attention 的模型结构如图 6-6 所示。

Model: "model\_2"

Layer (type)	Output Shape	Param #	Connected to
input_3 (InputLayer)	[(None, 60, 11)]	0	[]
bilstm (Bidirectional)	(None, 400)	339200	['input_3[0][0]']
dropout_4 (Dropout)	(None, 400)	0	['bilstm[0][0]']
attention_vec (Dense)	(None, 400)	160400	['dropout_4[0][0]']
dropout_5 (Dropout)	(None, 400)	0	['attention_vec[0][0]']
multiply_2 (Multiply)	(None, 400)	0	['dropout_4[0][0]', 'dropout_5[0][0]']
dense_2 (Dense)	(None, 1)	401	['multiply_2[0][0]']

Total params: 500,001  
Trainable params: 500,001  
Non-trainable params: 0

图 6-5 BiLSTM-Attention 网络结构 1

Model: "model"

Layer (type)	Output Shape	Param #	Connected to
input_1 (InputLayer)	[(None, 60, 11)]	0	[]
bilstm (Bidirectional)	(None, 60, 400)	339200	['input_1[0][0]']
dropout (Dropout)	(None, 60, 400)	0	['bilstm[0][0]']
bilstm2 (Bidirectional)	(None, 400)	961600	['dropout[0][0]']
attention_vec (Dense)	(None, 400)	160400	['bilstm2[0][0]']
dropout_1 (Dropout)	(None, 400)	0	['attention_vec[0][0]']
multiply (Multiply)	(None, 400)	0	['bilstm2[0][0]', 'dropout_1[0][0]']
dense (Dense)	(None, 1)	401	['multiply[0][0]']

Total params: 1,461,601  
Trainable params: 1,461,601  
Non-trainable params: 0

图 6-6 BiLSTM-Attention 网络结构 2

表 6-4 是比较了月粒度表现较好的 BiLSTM 模型，模型结构不同的 BiLSTM-Attention 模型的预测误差。比较发现，当 BiLSTM-Attention 模型使用 2 层 BiLSTM 层加上 1 层 Attention 层时，各评价指标的预测误差都最小。

表 6-4 BiLSTM 模型与 BiLSTM-Attention 模型的模型结构参数设置与预测误差

	BiLSTM 模型	BiLSTM-Attention 模型	
Attention 网络层数	0	1	1
BiLSTM 网络层数	2	1	2
滑动时间窗口	0	60	60
batch size	256	280	280
Epoch	50	35	66
mae	1252.534875	1283.175924	1196.399856
mape	40.72918063	43.42043008	36.99268951
mse	6780898.247	6796405.595	6768557.967
rmse	2604.015792	2606.991675	2601.645242

双层 BiLSTM 加单层 Attention 结构模型在验证集上的拟合效果如图 6-7 所示，尽管模型的误差有所减小，但它对验证集中需求量较大的情况拟合效果仍然不太好。

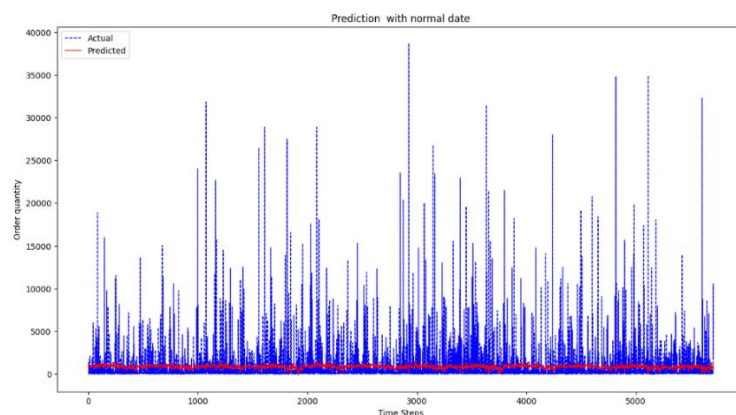


图 6-7 多变量 BiLSTM-Attention 模型的验证集真实值、预测值拟合效果

### 6.1.5 LightGBM 模型构建

本次实验用微软 LightGBM 框架实现对 LightGBM 模型的搭建。LightGBM 模型是基于决策树优化的，因此在构建模型前不再对数据做标准化处理，直接使用做了特征工程和排序的 month\_sum.csv 内数据。

模型选用了 GBDT 作为 Boosting model 来实现回归任务，由于模型可调整的超参数众多，使用了贝叶斯优化来替代手动调参实验。在实验中选择需要调参的模型，为多种超参数设置参数范围，调用 bayes\_opt 模块包来构建一个函数的后验分布（高斯过程），在搜索参数的过程中，改善后验分布，使算法找到设置的参数范围内的最佳超参数值。

实验中在设置的 LightGBM 的超参数范围以及优化后的最佳参数，如表 6-5 所示。

表 6-5 月粒度 LightGBM 参数类型和贝叶斯优化结果

参数	取值范围	最佳值
max_depth 树的最大深度	(5, 100)	19
num_leaves 叶子数量	(20, 200)	82
n_estimators 基学习器数量	(50, 1000)	227
learning_rate 学习率	(0.005, 0.5)	0.154654623
subsample 默认抽样	(0.5, 0.9)	0.65870699
colsample_bytree 特征抽样比率	(0.5, 0.99)	0.704340782
lambda_l1 L1 正则化系数	(0, 5)	3.601622467

lambda_l2 L2 正则化系数	(0, 3)	0.000343124
min_child_weight 最小叶子节点 样本权重	(1, 50)	5.524591144

贝叶斯优化后 LightGBM 模型预测误差如表 6-6 所示，模型在验证集上的预测值与真实值拟合效果如图 6-8 所示。

表 6-6 LightGBM 模型月粒度预测误差

损失评价指标	训练集	验证集
mac	655.1997731	681.9783096
mape	12.82534188	14.93459857
mse	2229394.585	2798085.526
rmse	1493.115731	1672.747897

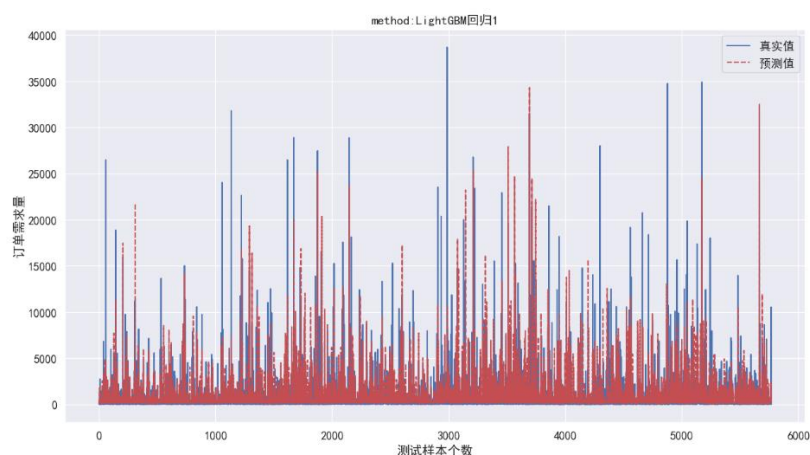


图 6-8 LightGBM 模型的月粒度验证集真实值、预测值拟合效果

### 6.1.6 预测精度对比与 LightGBM 模型选择

比较本节月粒度实验的几种模型在验证集中的拟合效果图（图 6-4，图 6-7，图 6-8），发现 LightGBM 模型的预测值和真实值的拟合效果最佳。

比较各模型的预测误差，可以找到按月粒度预测时，精度最好的模型。预测误差如表 6-7 所示，

每种模型的预测误差取优化完参数后该模型误差指标最小的结果。

表 6-7 月粒度各模型预测误差比较

损失评价指标	多变量			
	LightGBM	单变量 BiLSTM	多变量 BiLSTM	BiLSTM-Attention
mae	681.97831	1213.424675	1252.534875	1196.399856
mape	14.9345986	38.21214223	40.72918063	36.99268951
mse	2798085.53	6877186.311	6780898.247	6768557.967
rmse	1672.7479	2622.439	2604.015792	2601.645242

LightGBM 模型在月粒度下各损失评价指标都低于其他模型，其中 RMSE 指标比多变量 BiLSTM-Attention 低了约 35%。在月粒度下，LightGBM 模型对未来几月的需求量预测精度的优势明显。

## 6.2 月、周、天时间粒度预测精度对比

### 6.2.1 数据集处理

月时间粒度的数据集处理如上节（按月时间粒度预测）所示。

周时间粒度使用了通过特征工程整理的名为 `week_sum.csv` 的数据集，每个样本代表某商品在某地区某周的需求情况。天时间粒度使用了通过特征工程整理的名为 `day_sum.csv` 的数据集，每个样本代表某商品在某地区某天的需求情况。

对于数据集和验证集分割方法，按周、天时间粒度预测的分割方法与按月时间粒度一致。

对于预测误差的计算，按周/天时间粒度的方法会在各模型预测完验证集每周/天的需求量后，再把同地区同商品同一个月的需求量相加，统计验证集商品月需求量的真实和预测值，并以这组值来计算预测误差。

### 6.2.2 按周粒度模型预测精度对比

按与月粒度相同的方法和模型结构，构建了周粒度下的单变量 BiLSTM、多变量 BiLSTM 模型，这两个模型的超参数选用了在月粒度下多变量 BiLSTM 中表现最好的参数设置。当选用 MSE 和 RMSE 为主要损失评价指标时，单变量 BiLSTM 的预测精度略高于多变量 BiLSTM，因此在构建周

粒度 BiLSTM-Attention 模型时也选用单变量。

周粒度下的单变量 BiLSTM-Attention 模型结构与超参数设置与月粒度下表现最好的多变量 BiLSTM-Attention 相同，仅将输入模型的特征种类从多种变为历史产品需求量一种。

周粒度下的 LightGBM 模型构建方法、结构也与月粒度下的 LightGBM 相同，并且同样使用贝叶斯优化，搜寻得到周粒度下各超参数的最佳值。

周粒度下各模型的预测精度如表 6-8 所示，其中表现最佳的仍是 LightGBM 模型。

表 6-8 按周粒度各模型的验证集预测误差

损失评价指标	LightGBM	单变量 BiLSTM	多变量 BiLSTM	单变量 BiLSTM-Attention
mae	646.3619921	998.9759427	982.473307	1003.262013
mape	8.732199564	14.85485359	14.07029628	15.02269704
mse	2358078	6169452.427	6283374.642	6167558.996
rmse	1535.603464	2483.838245	2506.666041	2483.457066

### 6.2.3 按天粒度模型预测精度对比

使用与月粒度、周粒度相同的方法，针对天粒度对单/多变量 BiLSTM、单/多变量 BiLSTM-Attention、LightGBM 模型进行构建和预测实验。天粒度下各模型的预测精度如表 6-9 所示，当选取 RMSE 为主要损失评价指标，预测误差最小的模型是 LightGBM，其次是多变量 BiLSTM-Attention 模型。

表 6-9 按天粒度各模型的验证集预测误差

损失评价指标	LightGBM	单变量 BiLSTM	多变量 BiLSTM	单变量 BiLSTM-Attention	多变量 BiLSTM-Attention
mae	593.2660571	772.6663196	799.0290634	779.6322395	804.9643815
mape	6.248710378	5.714680185	6.439387462	5.89703661	6.621971294
mse	2283859.677	5201241.892	5018478.052	5023816.691	4986087.388
rmse	1511.244413	2280.623137	2240.195985	2241.387225	2232.954856

## 6.2.4 选取预测精度最高的粒度——天粒度

表 6-7、表 6-8、表 6-9 分别记录了月、周、天粒度下各模型的预测精度，每个粒度都统计了 BiLSTM、BiLSTM-Attention、LightGBM 模型预测误差即损失评价指标值，而模型的预测误差越小，预测精度越高。

比较可发现，同种模型的预测精度，从高到低排序，分为是按天、周、月粒度预测。

## 6.3 按天时间粒度的预测模型选择

### 6.3.1 LightGBM-BiLSTM-Attention 组合模型

在上一节中，比较得到天时间粒度下，预测误差最小的模型是 LightGBM，其次是多变量 BiLSTM-Attention 模型。考虑如下原因：BiLSTM-Attention 对处理序列数据具有一定优势；在天时间粒度下多变量 BiLSTM-Attention 模型的拟合效果（下图 6-9）相比月时间粒度（图 6-4）具有一定提升，在本节尝试组合 BiLSTM-Attention 和 LightGBM 模型。

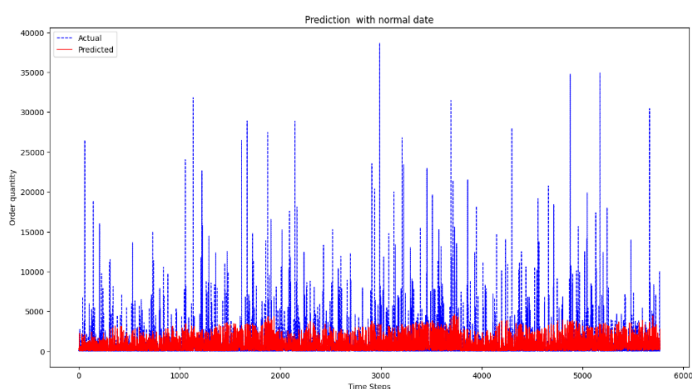


图 6-9 天时间粒度下多变量 BiLSTM-Attention 模型验证集拟合效果

组合模型的实现方法是：分别保存两个模型在验证集的预测值，两种预测值分别与对应的权重相乘，其结果再相加即为组合模型的预测值。再比较组合模型的预测值与验证集实际值，计算组合模型预测误差。

模型的权重采用前文介绍过的“预测均方根误差倒数法”，计算得到 LightGBM 和 BiLSTM-Attention 模型的参考权重分别约为 0.649142 和 0.350858。因此手动为组合模型分配了相近的权重，这些组合的权重和预测误差如表 6-10 所示：

表 6-10 LightGBM-BiLSTM-Attention 组合模型权重和预测误差



	组合 1	组合 2	组合 3	LightGBM 模型	BiLSTM-Attention 模型
LGBM 权重	0.6	0.65	0.7	1	0
BiLSTM-Attention 权重	0.4	0.35	0.3	0	1
mac	638.7677738	629.613397	621.0775104	593.2660571	804.9643815
mape	6.357899328	6.345427589	6.333260271	6.248710378	6.621971294
mse	2952107.743	2838208.4	2733705.609	2283859.677	4986087.388
rmse	1718.169882	1684.698311	1653.392152	1511.244413	2232.954856

### 6.3.2 选取 LightGBM 预测模型

比较表 6-10 组合模型与 LightGBM 模型的预测误差，发现组合模型的预测误差全部高于单 LightGBM 模型。并且 LightGBM 在天时间粒度下的拟合效果也极佳，如图 6-10 所示意。

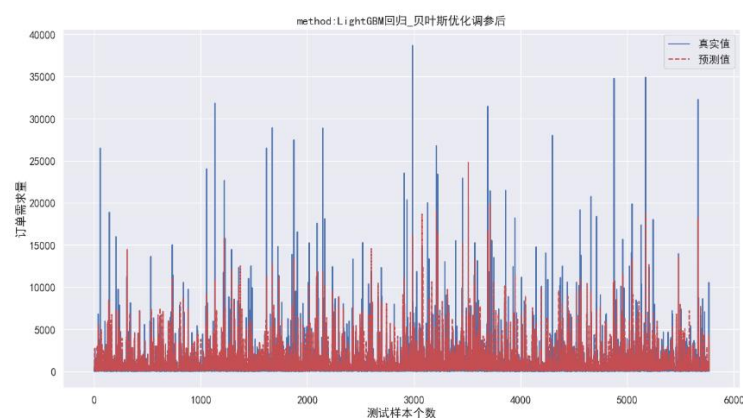


图 6-10 LightGBM 在天时间粒度下验证集拟合效果

因此本文选用 LightGBM 算法模型按天时间粒度的策略，来预测未来三月的商品月需求量，以取得较高的预测精度。天时间粒度下 LightGBM 模型的超参数设置，如表 6-11 所示。

表 6-11 天粒度 LightGBM 参数类型和贝叶斯优化结果

参数	取值范围	最佳值
max_depth 树的最大深度	(5, 100)	45

num_leaves 叶子数量	(20, 200)	145
n_estimators 基学习器数量	(50, 1000)	557
learning_rate 学习率	(0.005, 0.5)	0.053681683
subsample 默认抽样	(0.5, 0.9)	0.626206252
colsample_bytree 特征抽样比率	(0.5, 0.99)	0.519136844
lambda_l1 L1 正则化系数	(0, 5)	0.849152098
lambda_l2 L2 正则化系数	(0, 3)	2.63442751
min_child_weight 最小叶子节点 样本权重	(1, 50)	47.93658698

---

## 7 总结

通过对数据的分析，发现订单需求量受产品价格、地区、销售渠道、品类、时间等多个因素影响。首先，订单需求量受产品价格的影响，随着产品价格的增高，订单需求量会减少。但是在不同的销售环境中，产品价格与需求量之间的密切关系存在差异。例如，线下销售方式下产品价格与需求量之间的关系比在线上销售方式下更加密切，当产品在线下销售时，价格的变化对需求量的影响更为显著，需求量随着价格的升高而减少的趋势更为明显。在高端市场中，随着产品价格的增加，可能会吸引更多富裕的消费者，从而导致需求量的增加。而在低端市场中，价格的增加可能会导致需求量的减少。因此，企业需要根据不同的销售环境和目标消费者制定相应的定价策略，以实现最大化的销售收益。其次，不同地区的消费者对产品总需求量各不相同，这可能是由于地区的文化、经济、社会等因素不同所导致的。因此，企业需要根据不同地区的消费者需求量制定相应的销售策略，例如采取不同的定价、营销和产品推广策略，以满足不同地区消费者的需求并提高销售收益。第三，不同销售渠道、产品品类的需求量有不同的特点，而且在不同地区、不同产品品类等因素之间，销售渠道对需求量的影响也可能存在差异。总体上线上销售量高于线下销售量，企业可以加强线上销售渠道的营销力度，提高线上购物体验，使线上销售量再创新高。但是在某些情况下，线下销售渠道可能更适合特定的产品类型和消费者群体，而在一些情况下，线上销售渠道的需求量高于线下。例如，对于一些高端产品，线下销售渠道可能更能够提供个性化和专业化的服务，从而更能够满足高端消费者的需求。因此，企业需要结合不同产品类型和消费者群体等特点，灵活地选择和布局不同的销售渠道，以提高销售收益和满足不同消费者的需求。第四，需求量的变化可能会受到时间因素的影响，季节、时间段、节假日、促销活动等因素可能会对不同时间段的需求量产生影响。通过对数据的分析我们发现，秋季的订单需求量最高，其次是春季，而冬季的订单需求量；每月月中与每周周末的需求量高于平时最低；节假日、促销活动时消费者的购买意愿会增强。因此，了解这些时间因素对需求量的影响，可以帮助企业更好地规划和调整销售策略，以满足消费者需求并提高销售收益。根据不同时间段制定相应的销售和生产计划，避免过度生产和库存积压，同时保证供应充足。制定相应的促销计划，提高销售量，同时控制资金预算，避免过度的促销活动导致资金不足。

总之，通过对数据的分析，可以得出很多有价值的结论，从而为企业提供运营计划、资金预算、采购计划和生成计划等方面的指导。企业可以根据这些结论制定相应的措施，提高销售量，增强企业的竞争力。在数据分析的基础上，我们进行了特征工程。通过特征工程，构造了三类特征：时间戳衍生特征、时序值相关特征和属性变量相关特征来为后续的预测提供更多的信息。我们还进行了特征选择来去除冗余和无效特征，提高模型泛化能力，减少学习时间。

依据数据分析和特征工程，我们对产品月需求量进行预测，发现了相同算法模型按天时间粒度预测精度最高，按周时间粒度其次，按月时间粒度预测精度相比较而言最低。而在同一时间粒度中，LightGBM 算法模型相比单/多变量 BiLSTM 模型、BiLSTM-Attention 模型、LightGBM-BiLSTM-Attention 组合模型，预测误差指标更小，预测精度更高。本文采用贝叶斯优化调整 LightGBM 模型的结构和参数，取得了较好的效果。对于 BiLSTM 等模型的优化，使用了手动调整模型结构、Nadam

优化器和 EarlyStopping 早停机制，未来可考虑运用随机搜索等算法更好地优化模型参数。

## 参考文献

- [1] 胡鹏程. 基于时间序列的挖掘机销量预测算法及应用研究[D]. 南京: 东南大学, 2021.
- [2] CAI J, LUO J, WANG S, et al. Feature selection in machine learning: A new perspective[J]. *Neurocomputing*, 2018, 300: 70-79.
- [3] Graves A, Jürgen Schmidhuber. Framewise phoneme classification with bidirectional LSTM and other neural network architectures[J]. *Neural Networks*, 2005, 18(5-6): 602-610.
- [4] 李朝杨, 王希胤. 基于 Attention-BiLSTM 模型的 Python 源代码漏洞检测方法[J]. *华北理工大学学报(自然科学版)*, 2023, 45(02): 95-103+109.
- [5] P. Zhenbao, F. Shuhua, W. Hui. LightGBM technique and differential evolution algorithm-based multi-objective optimization design of DS-APMM. *IEEE Access*, 36 (2021), pp. 441-455, 10.1109/TEC.2020.3009480
- [6] 康艳, 程潇, 陈沛如, 等. 基于集成学习模型的非平稳月径流预报研究[J/OL]. *水资源保护*: 1-17[2022-05-27]
- [7] 季梦凡. 风速和风电场功率组合预测及其不确定性分析[D]. 北京: 华北电力大学, 2018.
- [8] 陈霄, 肖甫, 沙乐天等. 基于双层 BiLSTM 的安装程序 DLL 劫持漏洞挖掘方法[J/OL]. *软件学报*: 1-26[2023-04-28]. <https://doi.org/10.13328/j.cnki.jos.006763>.