### 分析流程 数据源： Q2.xlsx 算法配置： 算法： XGBoost回归 分析结果： XGBoost回归基于MSE、RMSE、MAE、MAPE、R²指标对模型进行评价，请看详细结论。

### 分析步骤 1. 通过训练集数据来建立XGBoost回归模型。 2. 通过建立的XGBoost来计算特征重要性。 3. 将建立的XGBoost回归模型应用到训练、测试数据，得到模型评估结果。 4. 由于XGBoost具有随机性，每次运算的结果不一样，若保存本次训练模型，后续可以直接上传数据代入到本次训练模型进行计算预测。 5. 注：XGBoost无法像传统模型一样得到确定的方程，通常通过测试数据预测精度来对模型进行评价。

### 详细结论

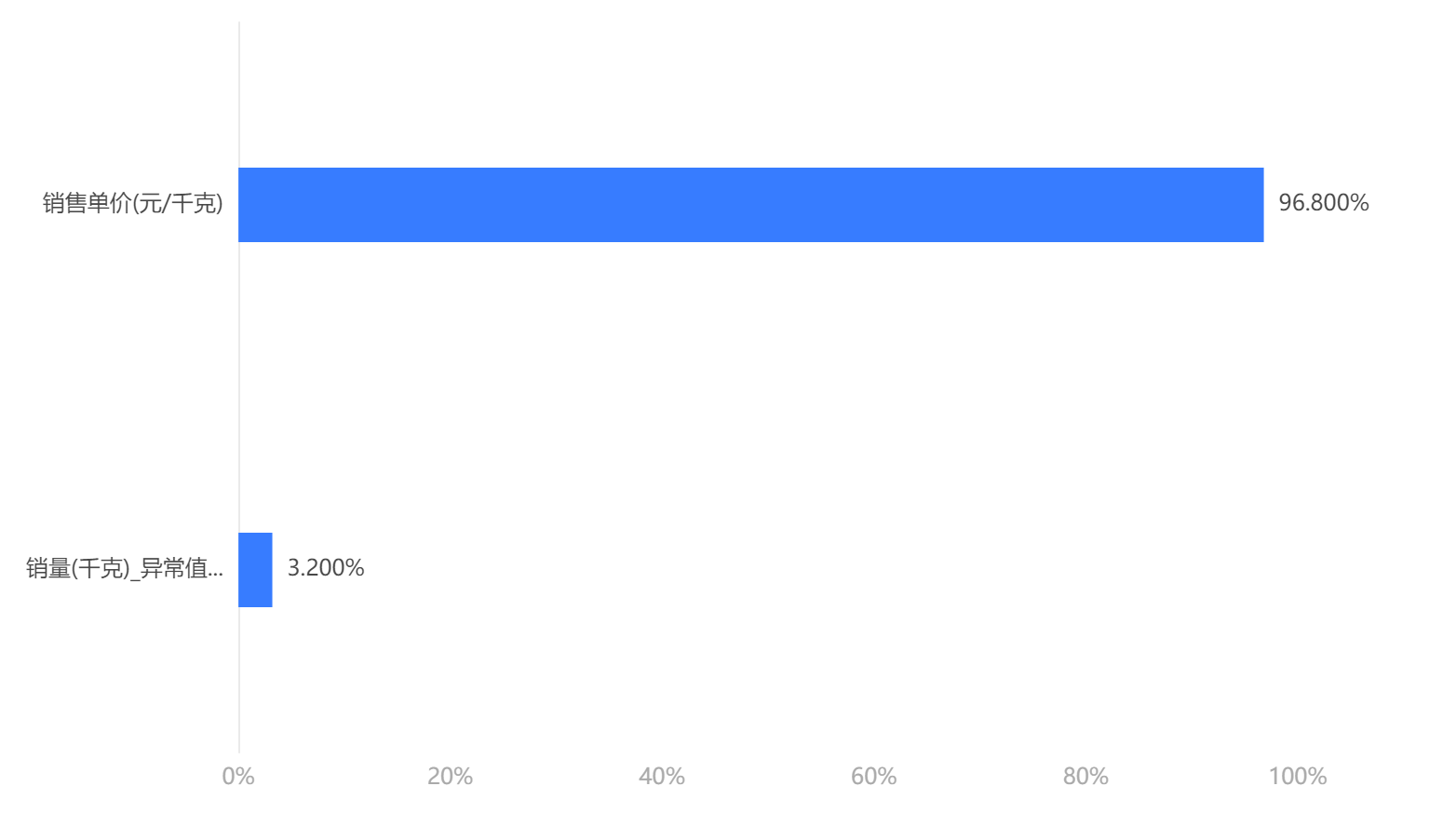
**输出结果1：模型参数**

|  |  |
| --- | --- |
| 参数名 | 参数值 |
| 训练用时 | 0.233s |
| 数据切分 | 0.8 |
| 数据洗牌 | 是 |
| 交叉验证 | 否 |
| 基学习器 | gbtree |
| 基学习器数量 | 120 |
| 学习率 | 0.0802 |
| L1正则项 | 0 |
| L2正则项 | 1 |
| 样本征采样率 | 1 |
| 树特征采样率 | 1 |
| 节点特征采样率 | 1 |
| 叶子节点中样本的最小权重 | 0 |
| 树的最大深度 | 10 |

**图表说明：**

上表展示了模型各项参数配置以及模型训练时长。

**输出结果2：特征重要性**



**图表说明：**

上柱形图或表格展示了各特征（自变量）的重要性比例。

**输出结果3：模型评估结果**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | MSE | RMSE | MAE | MAPE | R² |
| 训练集 | 0.111 | 0.333 | 0.224 | 5.936 | 0.988 |
| 测试集 | 1.004 | 1.002 | 0.692 | 20.694 | 0.855 |

**图表说明：**

上表中展示了交叉验证集、训练集和测试集的预测评价指标，通过量化指标来衡量XGBoost的预测效果。其中，通过交叉验证集的评价指标可以不断调整超参数，以得到可靠稳定的模型。  
● MSE（均方误差）： 预测值与实际值之差平方的期望值。取值越小，模型准确度越高。  
● RMSE（均方根误差）：为MSE的平方根，取值越小，模型准确度越高。  
● MAE（平均绝对误差）： 绝对误差的平均值，能反映预测值误差的实际情况。取值越小，模型准确度越高。  
● MAPE（平均绝对百分比误差）： 是 MAE 的变形，它是一个百分比值。取值越小，模型准确度越高。  
● R²： 将预测值跟只使用均值的情况下相比，结果越靠近 1 模型准确度越高。

**输出结果4：测试数据预测结果**

|  |  |  |  |
| --- | --- | --- | --- |
| 预测测试集结果Y | 利润(元/千克) | 销售单价(元/千克) | 销量(千克)\_异常值处理 |
| 12.554987907409668 | 13.27 | 29.8 | 60.052 |
| 0.978022575378418 | 0.8500000000000001 | 3 | 50.69 |
| 6.4318413734436035 | 5.6 | 18 | 62.87 |
| 4.67598295211792 | 6.36 | 14 | 48.096 |
| 4.287005424499512 | 3.18 | 10 | 44.776 |
| 1.0628148317337036 | 1.02 | 2.9 | 39.51 |
| 3.770261287689209 | 3.09 | 8 | 47.644 |
| 2.4184999465942383 | 2.69 | 6.8 | 45.674 |
| 1.7330260276794434 | 1.04 | 3 | 35.271 |
| 1.8424487113952637 | 2.1 | 5.8 | 115.423 |
| 5.0083818435668945 | 4 | 14 | 88.954 |
| 6.169584274291992 | 6.800000000000001 | 18 | 27.097 |
| 1.8374066352844238 | 1.81 | 4.5 | 160.777 |
| 8.319993019104004 | 8.400000000000002 | 21.6 | 11.525 |
| 7.500442981719971 | 7.200000000000001 | 21.6 | 74.574 |

**图表说明：**

上表格为预览结果，只显示部分数据，全部数据请点击下载按钮导出。  
上表展示了XGBoost对测试数据的预测情况。

**输出结果5：测试数据预测图**



**图表说明：**

图表说明：上图中展示了XGBoost对测试数据的预测情况。

**输出结果6：模型预测与应用**

**图表说明：**

● 系统会自动保存模型，需要注意的是：在机器学习中的XGBoost算法保存的模型是非常复杂的，不是类似于线性回归那样可以用一个公式保存，系统以二进制文件方式进行序列化保存。  
● 由于XGBoost具有随机性，每次训练的模型可能不一致，若保存本次训练模型，后续可以直接上传数据代入到本次训练模型进行计算预测。  
● 若删除本分析报告将会直接删除模型的缓存。

### 参考文献 [1] Scientific Platform Serving for Statistics Professional 2021. SPSSPRO. (Version 1.0.11)[Online Application Software]. Retrieved from https://www.spsspro.com. [2] Chen T , Guestrin C . XGBoost: A Scalable Tree Boosting System[J]. ACM, 2016.