### 分析流程 数据源： data\_因子分析.xlsx 算法配置： 算法： XGBoost回归 变量： 变量X:{Historical\_Achievement，Skill，Current\_Situation}；变量Y:{mark\_wave}收起 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

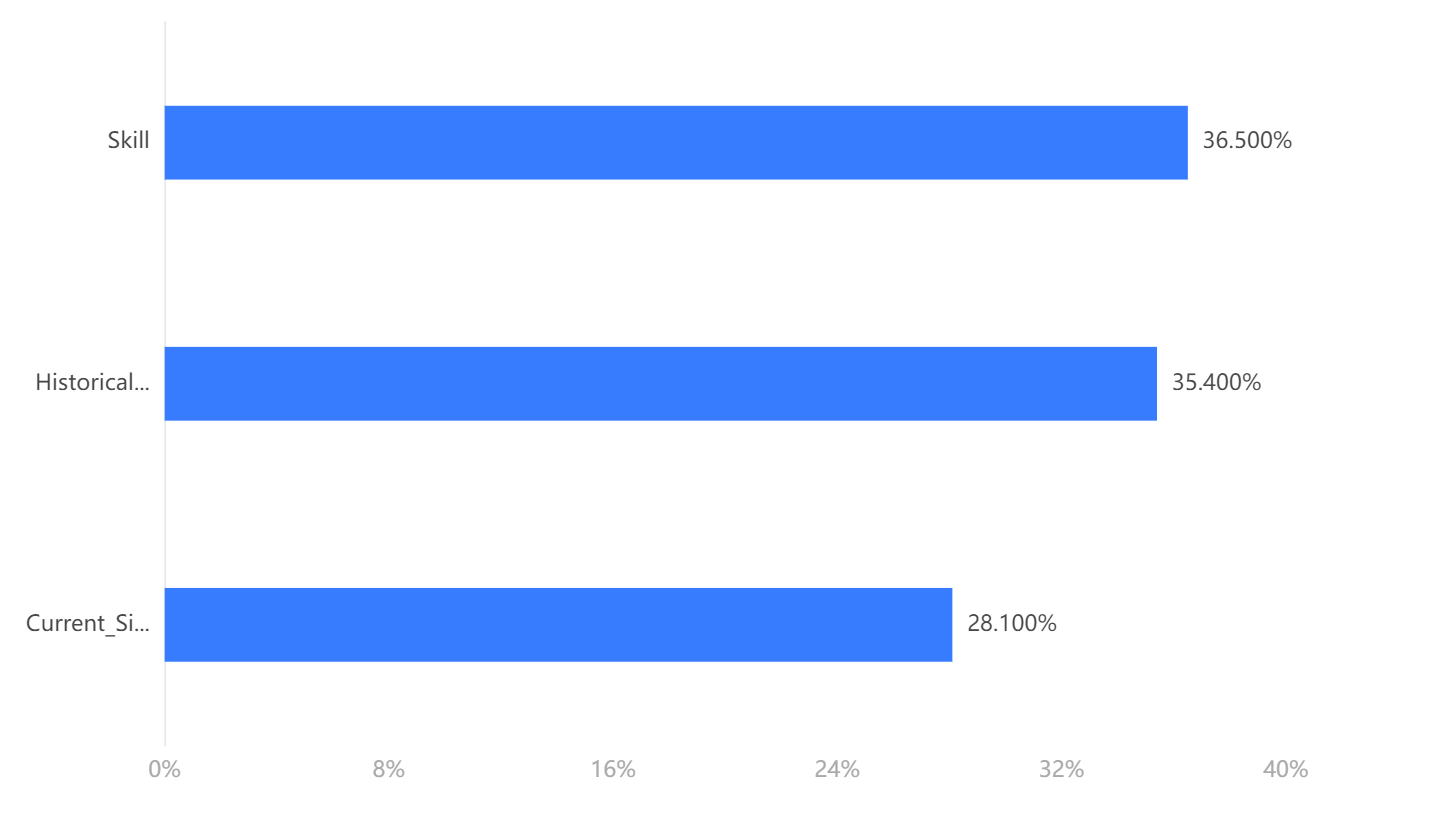
**输出结果1：模型参数**

|  |  |
| --- | --- |
| 参数名 | 参数值 |
| 训练用时 | 1.867s |
| 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 | 100 |
| 学习率 | 0.1 |
| 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 | 11.721 | 3.424 | 2.752 | 998.6 | 0.702 |
| 测试集 | 32.581 | 5.708 | 4.374 | 1019.097 | 0.231 |

**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。

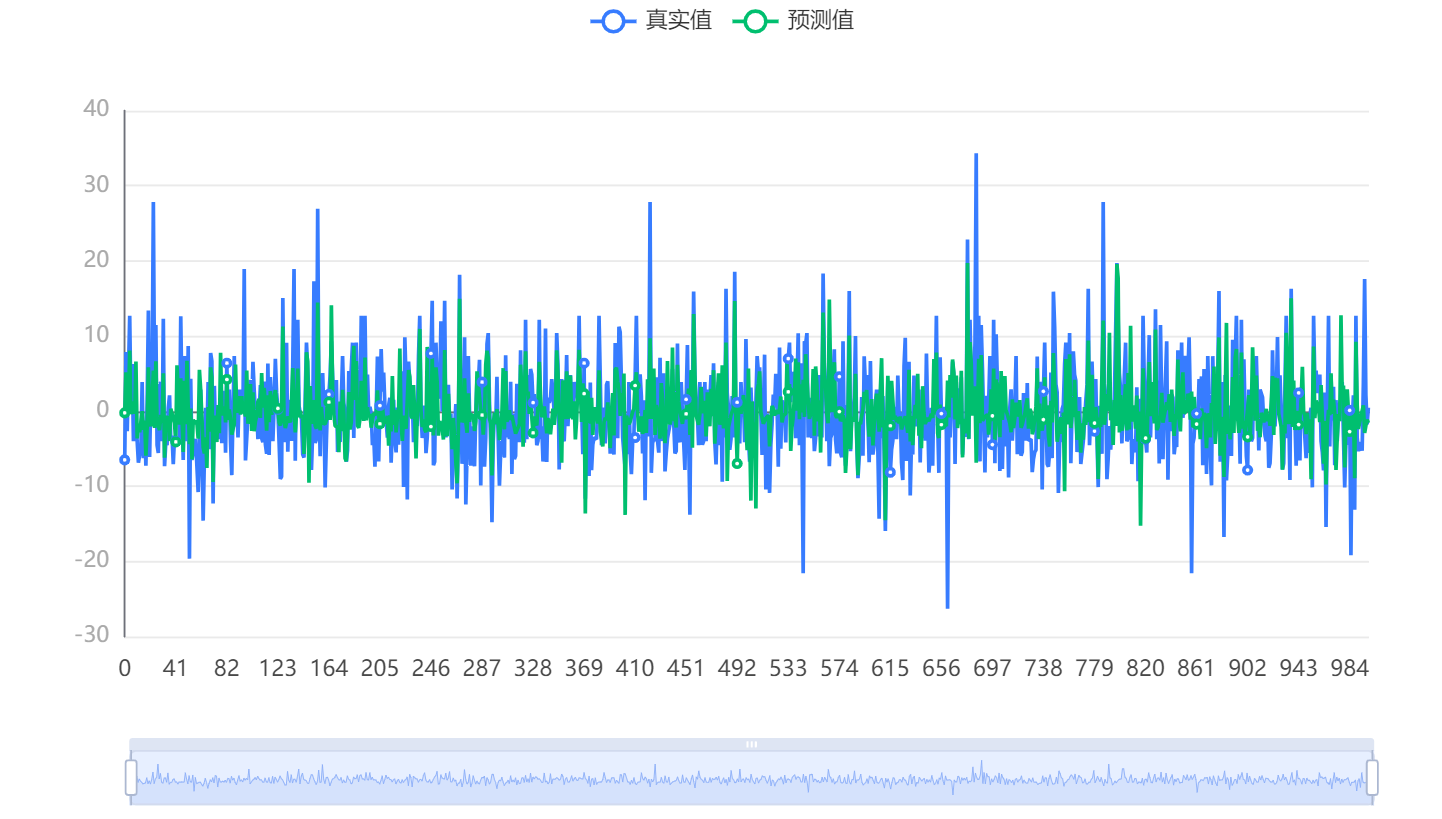
**输出结果4：预测结果**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 预测测试集结果Y | mark\_wave | Historical\_Achievement | Skill | Current\_Situation |
| -0.22495660185813904 | -6.482879409 | -0.5955075586316935 | 0.1402730327806191 | -0.3964108156463296 |
| 5.172699451446533 | 7.884177005 | 0.5681102166272617 | 0.02819128027049987 | 0.6038740423285981 |
| -0.10348767042160034 | -2.669732949 | -1.485968347866997 | -0.21974263435504968 | -0.03958668115825054 |
| 1.0895642042160034 | 5.060648212 | 0.38679998230414736 | -0.41519651889378195 | -0.20276344978149563 |
| 8.069219589233398 | 12.70759207 | 1.8050075945261945 | -0.09446901642460771 | -0.15788025997941244 |
| -0.388800710439682 | 0.466590919 | -0.5597045487234283 | 0.14069377596622268 | 0.14088078095473147 |
| 0.17512741684913635 | 6.272143018 | 1.1697841095567911 | -0.5084274763965371 | -0.21311397931230994 |
| 1.2792021036148071 | -4.005759563 | 0.21860950681245955 | -0.4180247634693935 | -0.8568246746174806 |
| -0.34160295128822327 | 1.716236279 | 0.6867771343061451 | -0.3894127168059673 | -0.8757914257994062 |
| 6.604124069213867 | -2.070226805 | 0.3882016925073235 | -0.4045917371279684 | 3.464115104703956 |
| -3.628833293914795 | -4.005759563 | 0.46244659594630605 | 4.61287465974025 | -0.776483936487876 |
| -1.759238600730896 | -6.865057165 | -1.1780858677467456 | 0.2156106758212432 | 0.1075943575487266 |
| -2.796658992767334 | -5.826788369 | -0.24744599620729626 | -0.34593173931264115 | -0.5571832835295994 |
| -3.0134053230285645 | -5.24581054 | -0.169296676298564 | -0.3309871842865 | -0.2121573326156611 |
| 0.0032069876324385405 | 3.878417441 | 0.02374538045471621 | 0.08528911683807588 | 0.6432542954809684 |

**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：**

上图中展示了XGBoost对测试数据的预测情况。预测图最多只展现测试集前1000个样本预测信息，若需要绘制完整的预测图，可在预测结果中导出数据重新绘图。

**输出结果6：模型预测与应用**

**图表说明：**

如果预测的数据超过15条，系统会自动展示前15条，其余数据请点击预览数据右上方下载按钮查看全部。

### 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.