### 分析流程 数据源： flights删除.csv 算法配置： 算法： CatBoost回归 分析结果： CatBoost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CatBoost回归模型。 2. 通过建立的CatBoost来计算特征重要性。 3. 将建立的CatBoost回归模型应用到训练、测试数据，得到模型评估结果。 4. 由于CatBoost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：CatBoost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

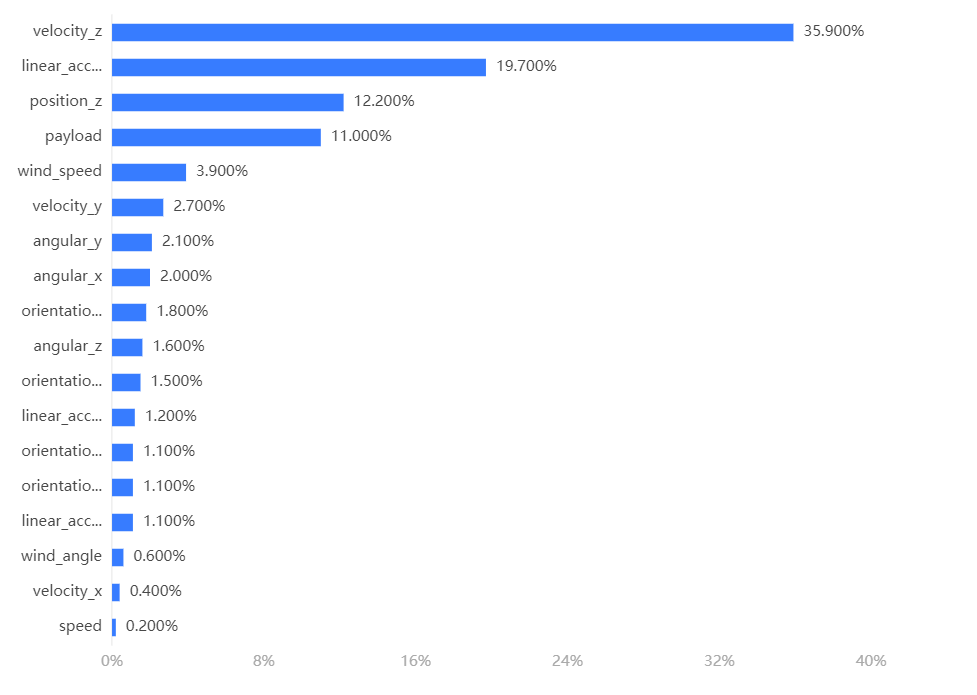
**输出结果1：模型参数**

|  |  |
| --- | --- |
| 参数名 | 参数值 |
| 训练用时 | 5.811s |
| 数据切分 | 0.7 |
| 数据洗牌 | 否 |
| 交叉验证 | 否 |
| 迭代次数 | 100 |
| 学习率 | 0.1 |
| L2正则项 | 1 |
| 树的最大深度 | 10 |
| 过拟合检测阈值 | 0 |
| 达成优化以后继续迭代的次数 | 20 |

**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 | 2193.007 | 46.83 | 30.137 | 1062.103 | 0.96 |
| 测试集 | 2790.356 | 52.824 | 33.262 | 68.914 | 0.948 |

**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CatBoost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。

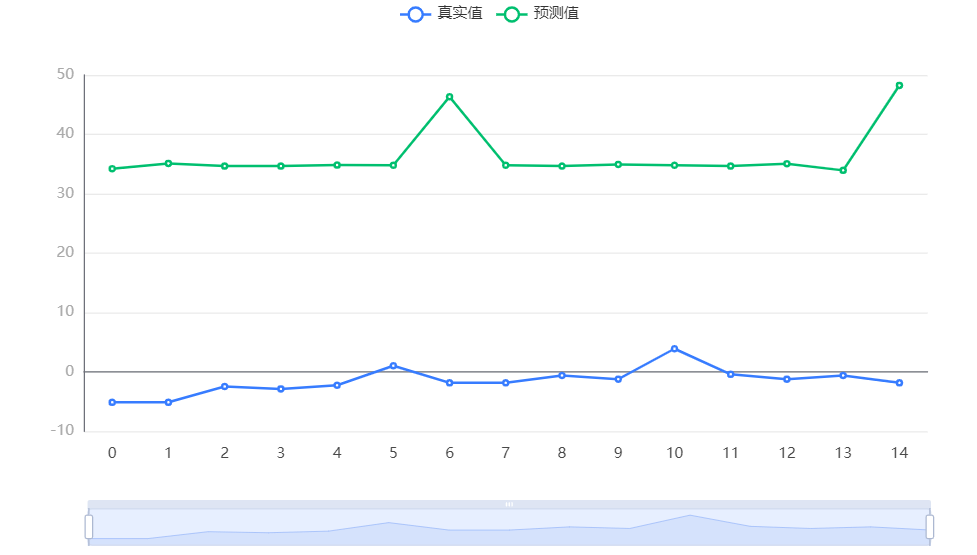
**输出结果4：预测结果**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 预测测试集结果Y | Power | wind\_speed | wind\_angle | position\_z | orientation\_x | orientation\_y | orientation\_z | orientation\_w | velocity\_x | velocity\_y | velocity\_z | angular\_x | angular\_y | angular\_z |
| 34.21155865141645 | -5.124614765 | 0.800000012 | 166 | 274.2107277 | 0.014172421 | 0.0001156 | -0.998996496 | 0.042487975 | -0.020923041 | -0.013161283 | 0.007608961 | 0.003926398 | 0.000140643 | -0.0000298 |
| 35.11387228985342 | -5.122872478 | 0.899999976 | 171 | 274.2115864 | 0.014162673 | 0.000121756 | -0.99899596 | 0.042503849 | -0.024353474 | -0.015231881 | 0.008192364 | -0.000904655 | 0.003742024 | 0.008428514 |
| 34.65645820350585 | -2.460890079 | 0.899999976 | 170 | 274.2197891 | 0.014151993 | 0.000122163 | -0.998997033 | 0.042483028 | -0.021577587 | -0.015419782 | 0.006983368 | 0.005417073 | -0.00000471 | 0.003597357 |
| 34.65645820350585 | -2.870620337 | 0.899999976 | 171 | 274.2267278 | 0.014179743 | 0.000136759 | -0.998996437 | 0.042485364 | -0.017069266 | -0.015981659 | 0.005742949 | -0.000586856 | 0.003161279 | 0.004333059 |
| 34.83291520331238 | -2.255049299 | 0.899999976 | 170 | 274.2273803 | 0.014181629 | 0.000122011 | -0.998996854 | 0.042476732 | -0.020841957 | -0.018440905 | 0.006369415 | -0.003962493 | -0.002275258 | -0.01062091 |
| 34.80643916534103 | 1.024873092 | 0.899999976 | 172 | 274.2340921 | 0.014196503 | 0.000158761 | -0.998995662 | 0.042499915 | -0.018193553 | -0.013103205 | 0.006295447 | -0.008643229 | -0.002455325 | -0.005049895 |
| 46.33009493451385 | -1.845488469 | 0.899999976 | 171 | 274.2399701 | 0.014229811 | 0.000182195 | -0.998995125 | 0.042499285 | -0.009962371 | -0.008060085 | 0.005869028 | 0.009060549 | 0.005904724 | 0.010248807 |
| 34.80643916534103 | -1.844592339 | 0.899999976 | 172 | 274.2406276 | 0.014222126 | 0.000184844 | -0.998994827 | 0.042509999 | -0.014505857 | -0.010256915 | 0.006564417 | -0.008494101 | -0.004974807 | -0.00436237 |
| 34.65645820350585 | -0.614983597 | 0.899999976 | 173 | 274.2454666 | 0.014240894 | 0.000185121 | -0.998995423 | 0.042490717 | -0.01810102 | -0.01461025 | 0.005501398 | 0.002282872 | 0.0000629 | 0.007455589 |
| 34.94115248866814 | -1.229907497 | 0.899999976 | 176 | 274.2461045 | 0.014233112 | 0.000173666 | -0.998995721 | 0.042485319 | -0.022144904 | -0.01691916 | 0.006219591 | -0.007826774 | -0.003067455 | -0.012196185 |
| 34.80643916534103 | 3.89451794 | 0.899999976 | 176 | 274.249683 | 0.014247415 | 0.000204951 | -0.998994887 | 0.042501055 | -0.018996501 | -0.011306951 | 0.005426459 | -0.005033327 | -0.001994395 | 0.001986743 |
| 34.65645820350585 | -0.410028893 | 0.899999976 | 175 | 274.2503082 | 0.014270074 | 0.000207839 | -0.99899447 | 0.042502653 | -0.02194571 | -0.013705932 | 0.006022733 | -0.002581591 | -0.00045339 | -0.008169486 |
| 35.05442780346624 | -1.22984771 | 0.800000012 | 174 | 274.2535146 | 0.014260859 | 0.000195828 | -0.998992741 | 0.04254603 | -0.02328883 | -0.015274044 | 0.005979096 | -0.005498691 | -0.001759588 | -0.004899241 |
| 33.93691749202213 | -0.614714781 | 0.800000012 | 175 | 274.2533619 | 0.0143163 | 0.000220675 | -0.998991966 | 0.042544924 | -0.014240259 | -0.010257382 | 0.001558982 | 0.002644413 | -0.000415108 | -0.003289259 |
| 48.2506602263 | -1.844592339 | 0.800000012 | 175 | 274.253594 | 0.014300313 | 0.000229559 | -0.998991728 | 0.042556383 | -0.017289023 | -0.012397502 | 0.002082918 | 0.004324922 | 0.006293554 | 0.006793559 |

**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CatBoost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CatBoost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：**

● 系统会自动保存模型，需要注意的是：在机器学习中的CatBoost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CatBoost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] Dorogush A V , Ershov V , Gulin A . CatBoost: gradient boosting with categorical features support[J]. 2018.