**《能源环境大数据案例分析》**

**课 程 实 验 报 告**

****

|  |  |
| --- | --- |
| **项目名称：** | **工业蒸汽量预测** |
| **学生姓名：** | **何郴** |
| **学生学号：** | **2103070062** |
| **学 期：** | **20 -20 学年第2学期** |
| **完成日期：** | **2023年3月15日** |

**实验一 工业蒸汽量预测**

# 一、实验背景与目的（10分）

## 1.1实验背景

火力发电的基本原理是：燃料在燃烧时加热水生成蒸汽，蒸汽压力推动汽轮机旋转，然后汽轮机带动发电机旋转，产生电能。在这一系列的能量转化中，影响发电效率的核心是锅炉的燃烧效率，即燃料燃烧加热水产生高温高压蒸汽。锅炉的燃烧效率的影响因素很多，包括锅炉的可调参数，如燃烧给量，一二次风，引风，返料风，给水水量；以及锅炉的工况，比如锅炉床温、床压，炉膛温度、压力，过热器的温度等。

## 1.2实验目的

本实验的目的是依据给定的脱敏后的锅炉传感器采集的数据，预测预测锅炉内蒸汽的产生量。

# 二．实验要求（10分）

预测结果以均方误差MSE（MeanSquaredE∏or ）作为评判标准计算公式如下：

其中其中为真实值为预测值。

# 三．实验原理描述（10分）

描述实验所用到的概念、原理及算法流程。

# 四．实验内容（20分）

## 4.1数据探索

### 4.4.1数据特征

表4.1 训练数据集

# Column Non-Null Count Dtype

--- ------ -------------- -----

0 V0 2888 non-null float64

1 V1 2888 non-null float64

2 V2 2888 non-null float64

3 V3 2888 non-null float64

4 V4 2888 non-null float64

5 V5 2888 non-null float64

6 V6 2888 non-null float64

7 V7 2888 non-null float64

8 V8 2888 non-null float64

9 V9 2888 non-null float64

10 V10 2888 non-null float64

11 V11 2888 non-null float64

12 V12 2888 non-null float64

13 V13 2888 non-null float64

14 V14 2888 non-null float64

15 V15 2888 non-null float64

16 V16 2888 non-null float64

17 V17 2888 non-null float64

18 V18 2888 non-null float64

19 V19 2888 non-null float64

20 V20 2888 non-null float64

21 V21 2888 non-null float64

22 V22 2888 non-null float64

23 V23 2888 non-null float64

24 V24 2888 non-null float64

25 V25 2888 non-null float64

26 V26 2888 non-null float64

27 V27 2888 non-null float64

28 V28 2888 non-null float64

29 V29 2888 non-null float64

30 V30 2888 non-null float64

31 V31 2888 non-null float64

32 V32 2888 non-null float64

33 V33 2888 non-null float64

34 V34 2888 non-null float64

35 V35 2888 non-null float64

36 V36 2888 non-null float64

37 V37 2888 non-null float64

38 target 2888 non-null float64

dtypes: float64(39)

该数据训练集含有2888个数据，其中V0-V37为特征变量，V38为标签值。且所以数据均无缺失值，数据类型为float64.

表4.2 测试数据集

# Column Non-Null Count Dtype

--- ------ -------------- -----

0 V0 1925 non-null float64

1 V1 1925 non-null float64

2 V2 1925 non-null float64

3 V3 1925 non-null float64

4 V4 1925 non-null float64

5 V5 1925 non-null float64

6 V6 1925 non-null float64

7 V7 1925 non-null float64

8 V8 1925 non-null float64

9 V9 1925 non-null float64

10 V10 1925 non-null float64

11 V11 1925 non-null float64

12 V12 1925 non-null float64

13 V13 1925 non-null float64

14 V14 1925 non-null float64

15 V15 1925 non-null float64

16 V16 1925 non-null float64

17 V17 1925 non-null float64

18 V18 1925 non-null float64

19 V19 1925 non-null float64

20 V20 1925 non-null float64

21 V21 1925 non-null float64

22 V22 1925 non-null float64

23 V23 1925 non-null float64

24 V24 1925 non-null float64

25 V25 1925 non-null float64

26 V26 1925 non-null float64

27 V27 1925 non-null float64

28 V28 1925 non-null float64

29 V29 1925 non-null float64

30 V30 1925 non-null float64

31 V31 1925 non-null float64

32 V32 1925 non-null float64

33 V33 1925 non-null float64

34 V34 1925 non-null float64

35 V35 1925 non-null float64

36 V36 1925 non-null float64

37 V37 1925 non-null float64

dtypes: float64(38)

该测试数据集，为1925个数据，其中V0-V37为特征变量，且没有缺失值，数据类型为float64。

### 4.1.2统计性描述

表4.3 训练数据集描述性统计

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **特征** | **V0** | **V1** | **V2** | **V3** | **V4** | **V5** | **V6** | **V7** | **V8** | **V9** | **V10** | **V11** | **V12** |
| **count** | 2886.00 | 2886.00 | 2886.00 | 2886.00 | 2886.00 | 2886.00 | 2886.00 | 2886.00 | 2886.00 | 2886.00 | 2886.00 | 2886.00 | 2886.00 |
| **mean** | 0.12 | 0.06 | 0.29 | -0.07 | 0.01 | -0.56 | 0.18 | 0.12 | 0.18 | -0.16 | 0.03 | -0.37 | 0.02 |
| **std** | 0.93 | 0.94 | 0.91 | 0.97 | 0.89 | 0.52 | 0.92 | 0.96 | 0.90 | 0.91 | 0.97 | 0.86 | 0.89 |
| **min** | -4.34 | -5.12 | -3.42 | -3.96 | -4.74 | -2.18 | -4.58 | -5.05 | -4.69 | -7.07 | -2.58 | -3.16 | -5.17 |
| **25%** | -0.29 | -0.22 | -0.31 | -0.65 | -0.39 | -0.85 | -0.31 | -0.30 | -0.16 | -0.39 | -0.42 | -0.80 | -0.42 |
| **50%** | 0.36 | 0.27 | 0.39 | -0.05 | 0.11 | -0.47 | 0.39 | 0.35 | 0.36 | 0.04 | 0.16 | -0.11 | 0.12 |
| **75%** | 0.73 | 0.60 | 0.92 | 0.62 | 0.55 | -0.15 | 0.83 | 0.78 | 0.73 | 0.04 | 0.62 | 0.25 | 0.62 |
| **max** | 2.12 | 1.92 | 2.83 | 2.46 | 2.69 | 0.49 | 1.90 | 1.92 | 2.25 | 1.34 | 4.83 | 1.46 | 2.12 |
| **特征** | **V13** | **V14** | **V15** | **V16** | **V17** | **V18** | **V19** | **V20** | **V21** | **V22** | **V23** | **V24** | **V25** |
| **count** | 2886.00 | 2886.00 | 2886.00 | 2886.00 | 2886.00 | 2886.00 | 2886.00 | 2886.00 | 2886.00 | 2886.00 | 2886.00 | 2886.00 | 2886.00 |
| **mean** | 0.20 | 0.02 | 0.10 | 0.11 | -0.04 | 0.06 | -0.11 | -0.19 | -0.06 | 0.30 | 0.16 | -0.02 | -0.05 |
| **std** | 0.92 | 1.02 | 1.03 | 0.98 | 0.66 | 0.95 | 1.11 | 0.79 | 0.78 | 0.64 | 0.97 | 1.03 | 0.92 |
| **min** | -3.68 | -2.46 | -2.90 | -5.98 | -2.22 | -3.58 | -3.70 | -3.40 | -2.64 | -1.38 | -5.54 | -1.34 | -3.81 |
| **25%** | -0.40 | -0.67 | -0.66 | -0.30 | -0.37 | -0.35 | -0.99 | -0.68 | -0.52 | -0.06 | 0.10 | -1.19 | -0.56 |
| **50%** | 0.29 | -0.16 | 0.00 | 0.31 | 0.17 | 0.08 | 0.00 | -0.16 | -0.06 | 0.21 | 0.34 | 0.10 | -0.08 |
| **75%** | 0.86 | 0.83 | 0.73 | 0.77 | 0.43 | 0.51 | 0.74 | 0.30 | 0.43 | 0.87 | 0.37 | 0.93 | 0.36 |
| **max** | 2.48 | 2.56 | 4.31 | 2.86 | 2.02 | 4.44 | 3.43 | 3.53 | 2.26 | 2.02 | 1.91 | 2.42 | 7.28 |
| **特征** | **V26** | **V27** | **V28** | **V29** | **V30** | **V31** | **V32** | **V33** | **V34** | **V35** | **V36** | **V37** | **target** |
| **count** | 2886.00 | 2886.00 | 2886.00 | 2886.00 | 2886.00 | 2886.00 | 2886.00 | 2886.00 | 2886.00 | 2886.00 | 2886.00 | 2886.00 | 2886.00 |
| **mean** | 0.07 | 0.27 | 0.14 | 0.10 | 0.06 | 0.13 | 0.02 | 0.01 | 0.01 | 0.20 | 0.03 | -0.13 | 0.13 |
| **std** | 0.89 | 0.27 | 0.93 | 1.06 | 0.89 | 0.87 | 0.90 | 1.01 | 1.00 | 0.99 | 0.97 | 1.02 | 0.98 |
| **min** | -5.13 | -1.16 | -2.44 | -2.91 | -4.51 | -5.86 | -4.05 | -4.63 | -4.79 | -5.70 | -2.61 | -3.63 | -3.04 |
| **25%** | -0.45 | 0.16 | -0.46 | -0.66 | -0.28 | -0.17 | -0.41 | -0.50 | -0.29 | -0.20 | -0.41 | -0.80 | -0.35 |
| **50%** | 0.08 | 0.33 | -0.45 | -0.02 | 0.05 | 0.30 | 0.04 | -0.04 | 0.16 | 0.36 | 0.14 | -0.19 | 0.31 |
| **75%** | 0.64 | 0.44 | 0.73 | 0.75 | 0.49 | 0.64 | 0.56 | 0.46 | 0.27 | 0.60 | 0.64 | 0.49 | 0.79 |
| **max** | 2.98 | 0.93 | 4.67 | 4.58 | 2.69 | 2.01 | 2.40 | 5.47 | 5.11 | 2.32 | 5.24 | 3.00 | 2.54 |

表4.3 测试数据集描述性统计

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **特征** | **V0** | **V1** | **V2** | **V3** | **V4** | **V5** | **V6** | **V7** | **V8** | **V9** | **V10** | **V11** | **V12** |
| **count** | 1925.00 | 1925.00 | 1925.00 | 1925.00 | 1925.00 | 1925.00 | 1925.00 | **1925.00** | 1925.00 | 1925.00 | 1925.00 | 1925.00 | 1925.00 |
| **mean** | -0.18 | -0.08 | -0.43 | 0.10 | -0.02 | 0.84 | -0.27 | **-0.17** | -0.27 | 0.26 | -0.05 | 0.55 | -0.03 |
| **std** | 1.07 | 1.08 | 0.97 | 1.03 | 1.15 | 0.96 | 1.05 | **1.04** | 1.09 | 1.01 | 1.04 | 0.95 | 1.14 |
| **min** | -4.81 | -5.49 | -4.28 | -3.28 | -4.92 | -1.17 | -5.65 | **-5.63** | -6.06 | -6.78 | -2.58 | -3.05 | -5.51 |
| **25%** | -0.66 | -0.45 | -0.98 | -0.64 | -0.50 | 0.12 | -0.73 | **-0.51** | -0.78 | -0.39 | -0.57 | -0.04 | -0.47 |
| **50%** | 0.07 | 0.20 | -0.27 | 0.22 | 0.12 | 0.44 | -0.08 | **0.02** | 0.00 | 0.40 | 0.08 | 0.43 | 0.13 |
| **75%** | 0.55 | 0.59 | 0.28 | 0.79 | 0.61 | 1.93 | 0.46 | **0.52** | 0.48 | 0.90 | 0.63 | 1.21 | 0.58 |
| **max** | 2.10 | 2.12 | 1.95 | 2.60 | 4.48 | 3.18 | 1.53 | **1.39** | 2.41 | 1.77 | 3.12 | 2.95 | 4.01 |
| **特征** | **V13** | **V14** | **V15** | **V16** | **V17** | **V18** | **V19** | **V20** | **V21** | **V22** | **V23** | **V24** | **V25** |
| **count** | 1925.00 | 1925.00 | 1925.00 | 1925.00 | 1925.00 | 1925.00 | 1925.00 | **1925.00** | 1925.00 | 1925.00 | 1925.00 | 1925.00 | 1925.00 |
| **mean** | -0.29 | -0.02 | -0.14 | -0.17 | 0.07 | -0.08 | 0.17 | **0.28** | 0.09 | -0.45 | -0.23 | 0.03 | 0.08 |
| **std** | 1.04 | 0.98 | 0.93 | 1.00 | 1.36 | 1.06 | 0.78 | **1.20** | 1.25 | 1.24 | 0.99 | 0.95 | 1.11 |
| **min** | -4.09 | -2.56 | -2.38 | -6.06 | -2.71 | -3.58 | -1.90 | **-2.86** | -9.18 | -2.51 | -5.55 | -1.34 | -4.73 |
| **100%** | -0.92 | -0.64 | -0.84 | -0.60 | -1.16 | -0.38 | -0.41 | **-0.57** | -0.96 | -1.38 | -0.22 | -0.63 | -0.54 |
| **125%** | -0.20 | 0.09 | -0.20 | -0.01 | -0.37 | 0.08 | 0.29 | **0.11** | 0.22 | -0.82 | -0.16 | 0.09 | -0.03 |
| **150%** | 0.41 | 0.58 | 0.49 | 0.46 | 1.23 | 0.44 | 0.75 | **0.97** | 1.02 | 0.31 | 0.36 | 0.80 | 0.46 |
| **max** | 2.59 | 2.86 | 3.07 | 2.03 | 3.09 | 4.44 | 2.20 | **4.06** | 4.34 | 3.61 | 1.35 | 1.80 | 6.62 |
| **特征** | **V26** | **V27** | **V28** | **V29** | **V30** | **V31** | **V32** | **V33** | **V34** | **V35** | **V36** | **V37** |  |
| **count** | 1925.00 | 1925.00 | 1925.00 | 1925.00 | 1925.00 | 1925.00 | 1925.00 | **1925.00** | 1925.00 | 1925.00 | 1925.00 | 1925.00 |  |
| **mean** | -0.11 | -0.41 | -0.21 | -0.15 | -0.08 | -0.19 | -0.03 | **-0.01** | -0.01 | -0.30 | -0.05 | 0.20 |  |
| **std** | 1.14 | 1.45 | 1.06 | 0.88 | 1.13 | 1.14 | 1.13 | **0.99** | 1.00 | 0.95 | 1.04 | 0.94 |  |
| **min** | -4.89 | -6.87 | -2.44 | -2.41 | -4.51 | -7.70 | -4.06 | **-4.63** | -4.79 | -7.48 | -2.61 | -3.35 |  |
| **175%** | -0.73 | -0.12 | -0.45 | -0.82 | -0.34 | -0.48 | -0.47 | **-0.46** | -0.29 | -0.35 | -0.59 | -0.43 |  |
| **200%** | 0.06 | 0.18 | -0.45 | -0.20 | 0.01 | 0.10 | 0.16 | **-0.04** | 0.16 | -0.27 | 0.08 | 0.15 |  |
| **225%** | 0.68 | 0.32 | -0.43 | 0.47 | 0.45 | 0.47 | 0.63 | **0.42** | 0.27 | 0.36 | 0.65 | 0.80 |  |
| **max** | 2.47 | 0.79 | 4.66 | 3.02 | 3.14 | 1.43 | 2.30 | **5.47** | 5.11 | 1.67 | 2.86 | 3.02 |  |

表4.4 数据表头探索

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **V0** | **V1** | **V2** | **V3** | **V4** | **V5** | **V6** | **V7** | **V8** | **V9** | **V10** | **V11** | **V12** |
| 0.566 | 0.016 | -0.143 | 0.407 | 0.452 | -0.901 | -1.812 | -2.36 | -0.436 | -2.114 | -0.94 | -0.307 | -0.073 |
| 0.968 | 0.437 | 0.066 | 0.566 | 0.194 | -0.893 | -1.566 | -2.36 | 0.332 | -2.114 | 0.188 | -0.455 | -0.134 |
| 1.013 | 0.568 | 0.235 | 0.37 | 0.112 | -0.797 | -1.367 | -2.36 | 0.396 | -2.114 | 0.874 | -0.051 | -0.072 |
| 0.733 | 0.368 | 0.283 | 0.165 | 0.599 | -0.679 | -1.2 | -2.086 | 0.403 | -2.114 | 0.011 | 0.102 | -0.014 |
| 0.684 | 0.638 | 0.26 | 0.209 | 0.337 | -0.454 | -1.073 | -2.086 | 0.314 | -2.114 | -0.251 | 0.57 | 0.199 |
| **V13** | **V14** | **V15** | **V16** | **V17** | **V18** | **V19** | **V20** | **V21** | **V22** | **V23** | **V24** | **V25** |
| 0.55 | -0.484 | 0 | -1.707 | -1.162 | -0.573 | -0.991 | 0.61 | -0.4 | -0.063 | 0.356 | 0.8 | -0.223 |
| 1.109 | -0.488 | 0 | -0.977 | -1.162 | -0.571 | -0.836 | 0.588 | -0.802 | -0.063 | 0.357 | 0.801 | -0.144 |
| 0.767 | -0.493 | -0.212 | -0.618 | -0.897 | -0.564 | -0.558 | 0.576 | -0.477 | -0.063 | 0.355 | 0.961 | -0.067 |
| 0.769 | -0.371 | -0.162 | -0.429 | -0.897 | -0.574 | -0.564 | 0.272 | -0.491 | -0.063 | 0.352 | 1.435 | 0.113 |
| -0.349 | -0.342 | -0.138 | -0.391 | -0.897 | -0.572 | -0.394 | 0.106 | 0.309 | -0.259 | 0.352 | 0.881 | 0.221 |
| **V26** | **V27** | **V28** | **V29** | **V30** | **V31** | **V32** | **V33** | **V34** | **V35** | **V36** | **V37** | **target** |
| 0.796 | 0.168 | -0.45 | 0.136 | 0.109 | -0.615 | 0.327 | -4.627 | -4.789 | -5.101 | -2.608 | -3.508 | 0.175 |
| 1.057 | 0.338 | 0.671 | -0.128 | 0.124 | 0.032 | 0.6 | -0.843 | 0.16 | 0.364 | -0.335 | -0.73 | 0.676 |
| 0.915 | 0.326 | 1.287 | -0.009 | 0.361 | 0.277 | -0.116 | -0.843 | 0.16 | 0.364 | 0.765 | -0.589 | 0.633 |
| 0.898 | 0.277 | 1.298 | 0.015 | 0.417 | 0.279 | 0.603 | -0.843 | -0.065 | 0.364 | 0.333 | -0.112 | 0.206 |
| 0.386 | 0.332 | 1.289 | 0.183 | 1.078 | 0.328 | 0.418 | -0.843 | -0.215 | 0.364 | -0.28 | -0.028 | 0.384 |

由以上表综合分析可知，38个特征值均为连续型特征。

## 4.2特征工程

## 4.3模型训练

## 4.4特征优化

## 4.5模型融合

# 五．实验结果及分析（20分）

截屏实验运行结果，并对实验结果进行分析。

# 六．实验体会（20分）

100字以上，不得抄袭。

# 七．附程序代码清单（10分）

可上传附件。

注：1利用公式编辑器输入公式

2正文中西文字体要求：小四号宋体和**Times New Roman**