机器学习数据集模板：工程名称、资质要求、最高报价、下浮率、工资、开标时间、镇区、计划工期、投标人数、P值下浮率

- 工程名称量化，采用词频分析
- 资质要求，按序号
- 开标时间量化，按数字
- 镇区量化，按序号

```
clear;clc;close all;
% 招标文件汇总
zb_pdf = readtable('\\longde\longde\008-投标中心\006-数据分析中心\招标文件PDF汇总\招标文件汇总.xlsx',
                   "VariableNamingRule", "preserve");
% 投标项目
zb_xlsx = GetDirNames('\\longde\longde\008-投标中心\006-数据分析中心\投标数据Excel文件汇总\*.xlsx'
% 交集
[~, ~, indx] = intersect(zb_xlsx, zb_pdf.("工程名称"));
zb_pdf = zb_pdf(indx, :);
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%% 提取数据
len = size(zb_pdf, 1);
project_names= zb_pdf.("工程名称");
qualification = cellfun(@GetQualificationNumeric, zb_pdf.("资质要求"));
Ms = zb_pdf.("最高报价");
R = zb_pdf.("下浮率");
Salary = zb_pdf.("工资");
bid_time = cellfun(@GetBidTimeNumeric, zb_pdf.("开标时间"));
project_address = cellfun(@GetDistrictAndTownNumeric, zb_pdf.("镇区"));
bid_company_count = nan(len, 1);
project_date = zb_pdf.("计划工期");
Rp = nan(len, 1);
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%% 开始遍历计算
tic;
parpool("local");
```

```
Starting parallel pool (parpool) using the 'local' profile ...
Connected to the parallel pool (number of workers: 4).
```

```
parfor i = 1:len
    filename = ['\\longde\longde\008-投标中心\006-数据分析中心\投标数据Excel文件汇总\', ...
                project_names{i}, '.xlsx'];
    try
        file = readtable(filename, 'Sheet', 1, 'VariableNamingRule', 'preserve');
        Gn = file.("投标报价");
    catch ME
        file = readtable(filename, 'Sheet',  '开标信息', 'VariableNamingRule', 'preserve');
        Gn = file.("投标报价");
    end
    % 去掉空白值
    index = isnan(Gn);
    Gn(index) = [];
    bid_count = length(Gn);
    Gn_sort = sort(Gn);
    if bid_count > 19
```

```
        % 采用去尾法取整
        cut = floor(0.2 * bid_count);
    elseif bid_count > 4
        % 首尾去掉 1 个
        cut = 1;
    else
        cut = 0;
    end
    % 去最高最低平均价 P
    Gn_cal = Gn_sort(cut+1 : bid_count-cut);
    P = mean(Gn_cal);
    % 获取两个关键字段
    bid_company_count(i) = bid_count;
    Rp(i) = 1 - P / Ms(i);
    % 打印进度
    if mod(i, 10) == 0
        sprintf('进度:  %d / %d \n', i, len);
    end
end
```

```
进度:  10 / 366
进度:  20 / 366
进度:  60 / 366
进度:  190 / 366
进度:  100 / 366
进度:  140 / 366
进度:  50 / 366
进度:  90 / 366
进度:  180 / 366
进度:  130 / 366
进度:  40 / 366
进度:  80 / 366
进度:  170 / 366
进度:  30 / 366
进度:  120 / 366
进度:  160 / 366
进度:  70 / 366
进度:  110 / 366
进度:  150 / 366
进度:  210 / 366
进度:  280 / 366
进度:  250 / 366
进度:  230 / 366
进度:  200 / 366
进度:  270 / 366
进度:  240 / 366
进度:  220 / 366
进度:  300 / 366
进度:  290 / 366
进度:  260 / 366
进度:  310 / 366
进度:  320 / 366
进度:  330 / 366
进度:  340 / 366
进度:  360 / 366
进度:  350 / 366
```

```
delete(gcp('nocreate'));
toc;
```

历时 60.231491 秒。

```
% 保存
names = {'工程名称', '资质要求', '最高报价', '下浮率', '工资', '开标时间', '镇区', ...
         '计划工期', '投标人数', 'P值下浮率'};
result = table(project_names, qualification, Ms, R, Salary, bid_time, project_address, ...
               project_date, bid_company_count, Rp, 'VariableNames', names);
zb_ML = '\\longde\longde\008-投标中心\006-数据分析中心\分析结果汇总\zb_ML.xlsx';
writetable(result, zb_ML);
winopen(zb_ML);
```