2010-2018年样本信息

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 年份 | 样本总数 | AKI\_0样本数 | AKI\_1样本数 | AKI\_2样本数 | AKI\_3样本数 |
| 2010 | 15081 | 12452 | 2261 | 214 | 154 |
| 2011 | 15568 | 12971 | 2253 | 189 | 155 |
| 2012 | 16816 | 14135 | 2308 | 196 | 177 |
| 2013 | 17586 | 15162 | 2041 | 195 | 188 |
| 2014 | 18862 | 16213 | 2218 | 240 | 191 |
| 2015 | 20265 | 17482 | 2294 | 237 | 222 |
| 2016 | 20566 | 17698 | 2409 | 237 | 222 |
| 2017 | 18167 | 15742 | 2022 | 195 | 208 |
| 2018 | 2410 | 2126 | 244 | 26 | 14 |

2010-2018年将样本根据label特征进行拆分成多个样本之后的信息label[[1,2],[2,3]]

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 年份 | 样本总数 | AKI\_0样本数 | AKI\_1样本数 | AKI\_2样本数 | AKI\_3样本数 |
| 2010 | 15323(15081) | 12452(12452) | 2455(2261) | 262(214) | 154(154) |
| 2011 | 15815(15568) | 12971(12971) | 2441(2253) | 248(189) | 155(155) |
| 2012 | 17071(16816) | 14135(14135) | 2496(2308) | 263(196) | 177(177) |
| 2013 | 17856(17586) | 15162(15162) | 2238(2041) | 268(195) | 188(188) |
| 2014 | 19190(18862) | 16213(16213) | 2472(2218) | 314(240) | 191(191) |
| 2015 | 20621(20265) | 17482(17482) | 2565(2294) | 322(237) | 252(222) |
| 2016 | 20902(20566) | 17698(17698) | 2659(2409) | 323(237) | 222(222) |
| 2017 | 18471(18167) | 15742(15742) | 2245(2022) | 276(195) | 208(208) |
| 2018 | 2434(2410) | 2126(2126) | 264(244) | 30(26) | 14(14) |

Demo: 人口统计学信息 一维[ ] 【元素个数固定（4个）】

[

'70', ---Age

'N', ---Hispanic

'03', ---Race

'M' ---Sex

]

Vital: 生命体征 三维[ [ ] ] 【元素个数固定（8个），与时间有关的数值个数不固定】

[

[ --Height [ 数值，时间 ]

[ '67', '0.96' ],

[ '65', '8.96' ], ...

],

[ --Weight [ 数值，时间 ]

[ '130', '0.96' ],

[ '120', '8.96' ], ...

],

[ --BMI [ 数值，时间 ]

[ '20', '0.96' ],

[ '20', '8.96' ], ...

],

[ --Smoking [ 数值，时间 ]

[ '03', '0.96' ],

[ '03', '1.96' ],

[ '01', '8.96' ], ...

],

[ --Tabacco [ 数值，时间 ]

[ '02', '0.96' ],

[ '02', '1.96' ],

[ '02', '8.96' ], ...

],

[ --TabaccoType [ 数值，时间 ]

[ '04', '0.96' ],

[ '04', '1.96' ],

[ '01', '8.96' ], ...

],

[ --SBP [ 数值，时间 ]

[ '112', '0.67' ],

[ '110', '0.71' ],

[ '100', '0.79' ],

[ '100', '0.83' ], ...

],

[ --DBP [ 数值，时间 ]

[ '68', '0.67' ],

[ '71', '0.71' ],

[ '60', '0.79' ],

[ '74', '0.83' ], ....

]

]

Lab: 实验室测量值 四维[ [ [ [ ] ] ] ] 【元素个数不固定（817个）】

[

[

[ --lab index

[ 'lab108' ]

],

[ --[ 数值，类型，时间 ]

[ '8.8', 'MG/DL', '1' ],

[ '8.9', 'MG/DL', '2' ],

[ '8.4', 'MG/DL', '3' ], ...

]

] ,

[

[

[ 'lab140' ]

],

[

[ '25', 'MMOL/L', '1' ], ...

]

], ...

]

Css: 入院诊断 三维[ [ [ ] ] ] 【元素个数不固定（280），特别关注css26,css2601,...css2621】

[

[

[ 'ccs50' ], --css index

[ --时间

'-350',

'-340', ...

]

],

[

[ 'ccs133' ],

[

'-197', ...

]

], ...

]

或 [[['0']]] 表示该特征没有数据【数据缺失】

Px：合并症 三维[ [ [ ] ] ] 【元素个数不固定（15606个）】

[

[

[ 'px11163' ], --px index

[ --时间

'1',

'2',

'3', ...

]

], ...

[

[ 'px13069' ],

[

'1', ...

]

], ...

]

Med：入院前及住院期间的药物、病史等 四维[ [ [ [ ] ] ] ] 【元素个数不固定（11538个）】

[

[

[ --med index

[ 'med1004' ]

],

[ --[ 数值（日常接触），时间 ]

[ '1', '1' ],

[ '1', '2' ],

[ '1', '3' ], ...

]

],

[

[

[ 'med10670' ]

],

[

[ '1', '1' ],

[ '1', '2' ],

[ '1', '3' ],

[ '1', '4' ],

[ '1', '5' ], ...

]

], ...

]

Label：AKI状态 二维[ [ ] ] 【元素个数不固定（4个/或1个）】

[ ---[ 状态，时间 ]

[ '1', '2' ],

[ '2', '3' ],

[ '2', '7' ], ...

]

数据预处理中的变化率：

Vital:（Height,weight,BPSYS,BMI,BPDIA）

mean：平均值

nearValue：临近值

scale: 比例值（最大值 - 最小值）

rate：变化率：（最大值-最小值）/（最大值所对应时间 - 最小值所对应时间）

vital\_SMOKE\_elev：吸烟趋势（-1：减轻了,0：没变,1：加重了，2：未知，-1000：缺失）

Lab:

mean；

nearValue；

scale

CCS，Med:

count：该标签下有多少记录数，eg：[ccs1：-300，-286，-200]; 则ccs1\_count = 3

nearValue;

AKI：

AKI\_0,

AKI\_1,

AKI\_2,

AKI\_3,

AKI\_former：当前测试时间之前的AKI状态记录