



基于RDNN的睡眠分析与疾病预测模型



RDNN-based Sleep Analysis and Disease Prediction Mathematical Model

汇报人：0921队

导师：朱佳

目录

CONTENTS

01

题目背景

Problem Background

02

研究方法

Research Methods

03

模型建立

Model Establishment

04

问题求解

Problem Solving

05

评估对比

Evaluation and Comparison

06

模型优缺点

Advantages and Disadvantages

PART ONE

问题 背景

问题背景

PART ONE

睡眠质量与人体各项参数指标有什么关系？

PART TWO

睡眠质量的好坏是否可以反映身体潜在的疾病？

PART THREE

人的身体指标的检测需要花费较大的成本，是否可以通过睡眠指标来预测疾病？

问题背景



PART TWO

研究 方法

研究方法

第一步

数据统计与可视化

利用统计学的方法，绘制不同睡眠质量下同一睡眠指标分布**折线图**和**热度图**，并计算相应的**方差**。

数据预处理

对附件II进行**数据清洗**，去除干扰和无效数据，对剩余数据**统计并分类**。

第二步

第三步

建立RDNN模型

构建再分类神经网络Reclassified Deep Neural Networks(RDNN)模型，对睡眠质量数据分类病情进行建模，得到最终的病情预测。

设计睡眠改善程序

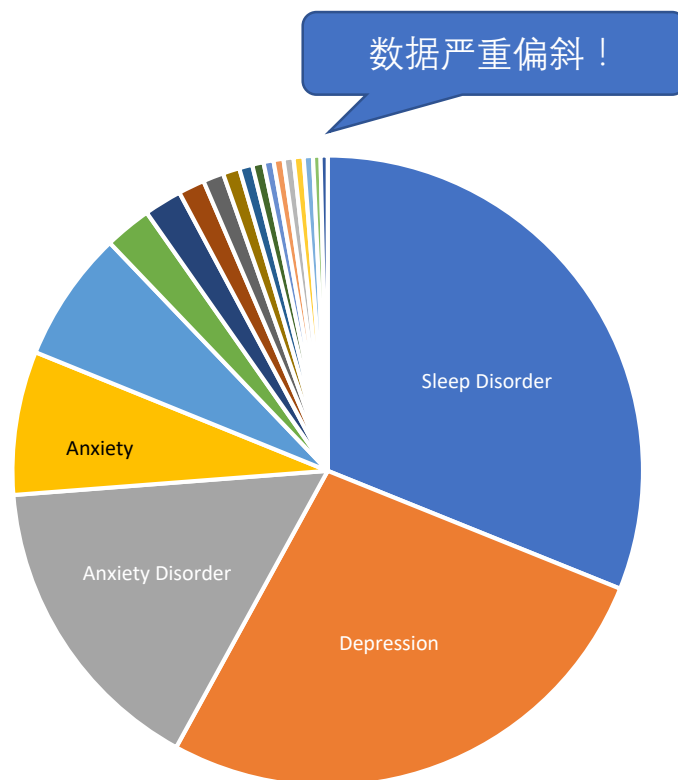
设计问卷收集用户的睡眠信息，合理**量化**为RDNN能处理的指标。

第四步

模型 建立

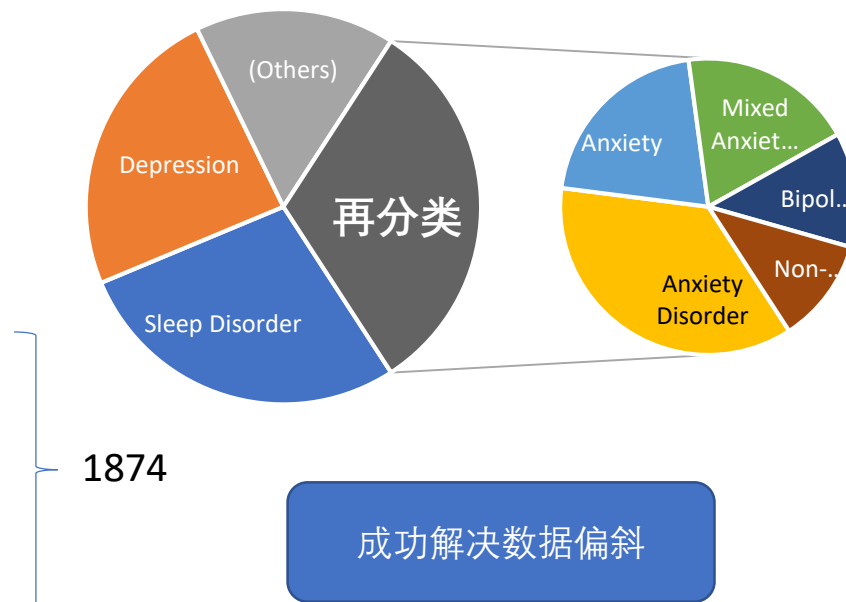
数据预处理之精髓：再分类

疾病名称	患者数量
Sleep Disorder	1703
Depression	1472
Anxiety Disorder	865
Anxiety	403
Mixed Anxiety And Depression	368
Bipolar Affective Disorder	132
Non-Organic Insomnia	106
Recurrent Depressive Disorder	74
Adjustment Disorder	58
Schizophrenia	47
Consultation	37
...	...

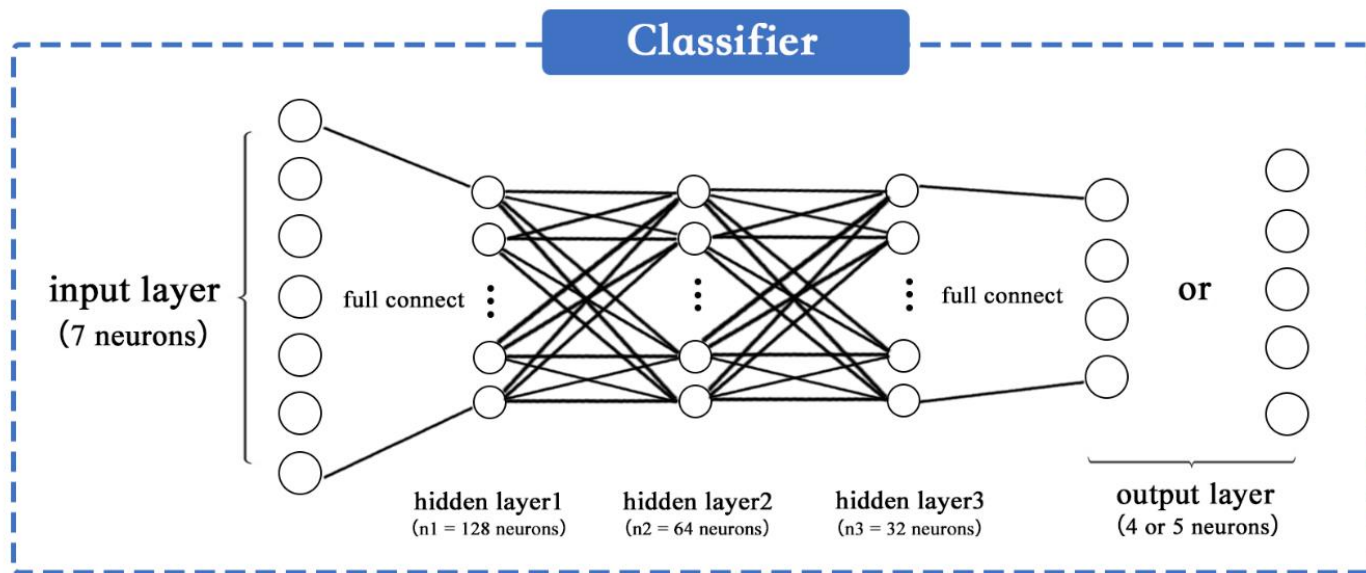


数据预处理之精髓：再分类

疾病名称	患者数量
Sleep Disorder	1703
Depression	1472
Anxiety Disorder	865
Anxiety	403
Mixed Anxiety And Depression	368
Bipolar Affective Disorder	132
Non-Organic Insomnia	106
(Others)	~1000



RDNN Structure



$$L(R) = - \sum_{n=1}^N R(x) * \log\{f[I(x)]\}$$

N - 训练集样本

$I(x)$ - 为输入数据

$R(x)$ - (网络所学) 监督标签

$f(x)$ - 分类模型

$L(x)$ - 损失函数

More specific RDNN

Our loss function:

$$L(R) = - \sum_{n=1}^N R(x) \log\{f[I(x)]\}$$

Neuron Layer:

$I^0 = \text{input layer}$

$I^l = \sigma_1(W^l * I^{l-1} + b^l) \quad \text{Layer 1, 2, 3}$

$O = \sigma_2(W^l * I^{l-1} + b^l) \quad \text{Layer 4}$

σ_1 - ReLU激活函数

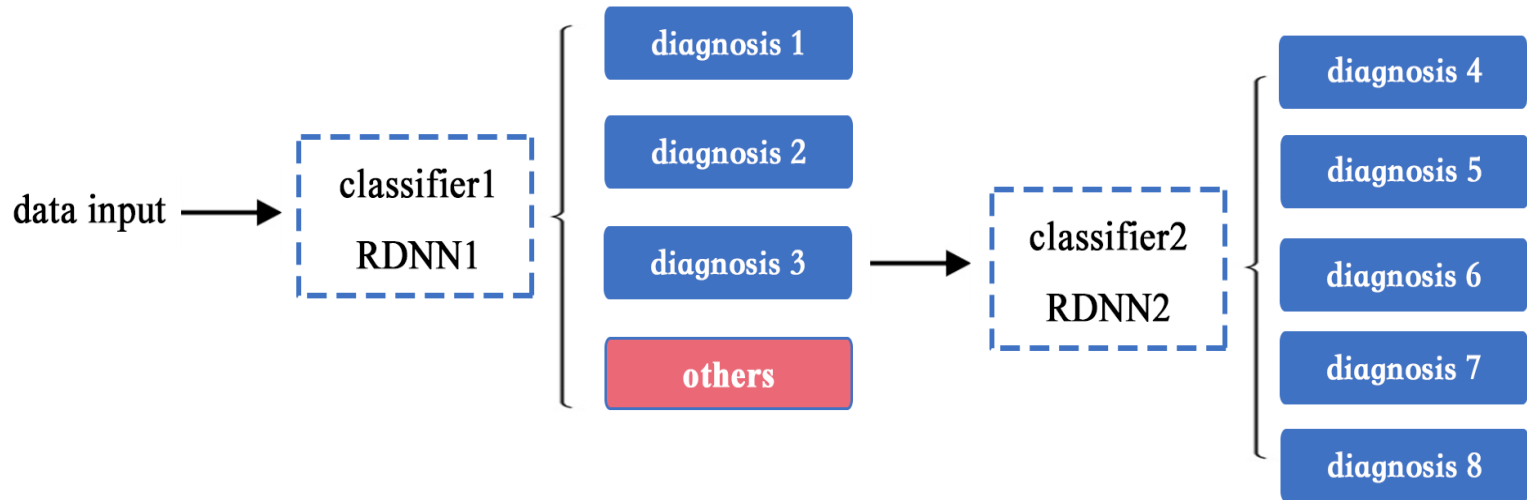
σ_2 - SoftMax激活函数

I - 输入层

O - 输出层

Classifier architecture

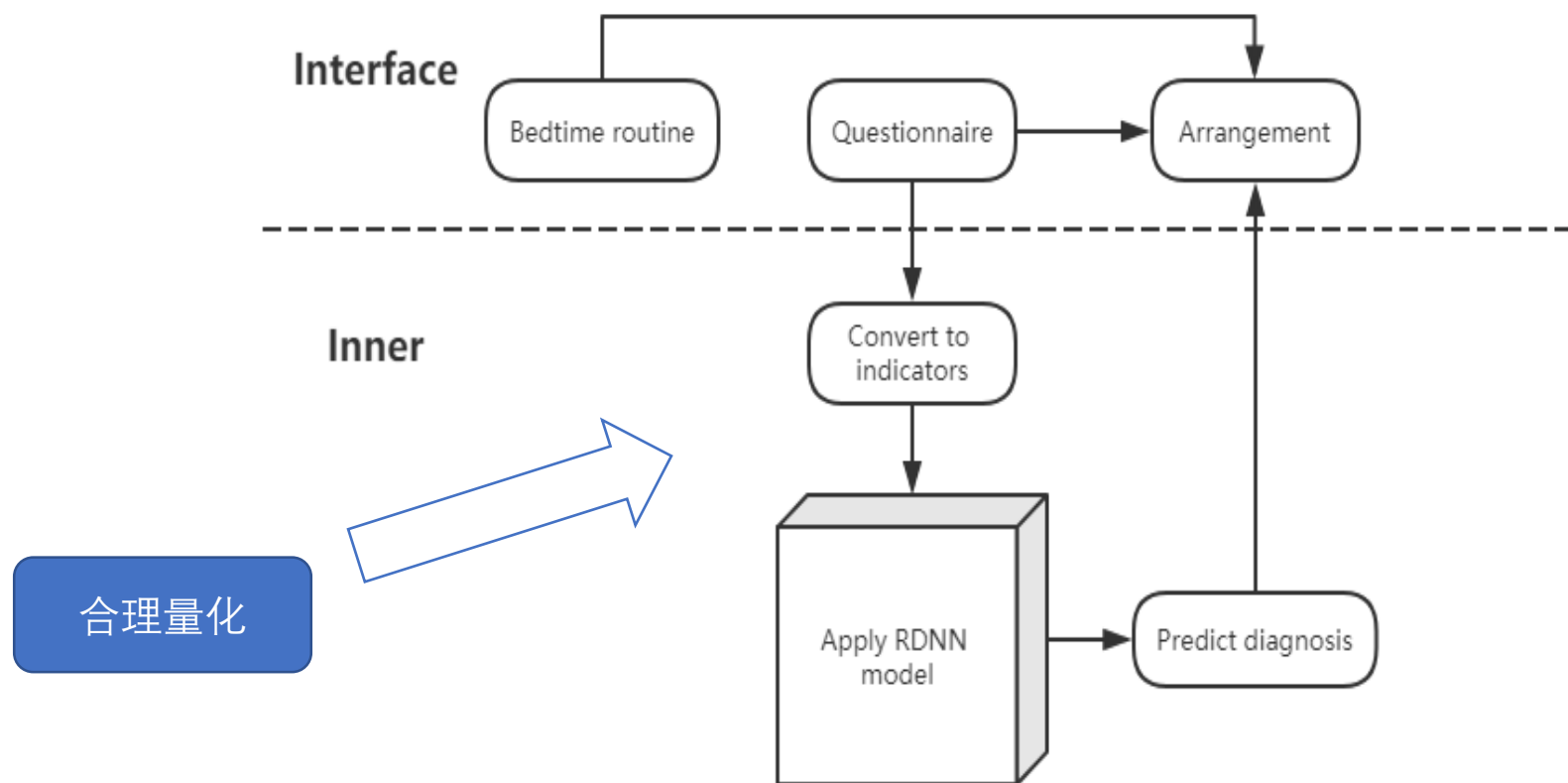
Multi-label precision classification



Two stage classifier to solve the data skew problem

Four neuron layer to classify the output result precisely

睡眠诊断程序结构



将问卷结果量化为RDNN模型能够接收的数据，将数据放入模型中预测疾病状况。

问卷设计的合理量化



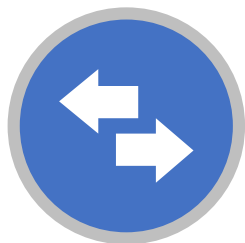
以详细时长为选项

对于睡眠时长和睡眠延迟，我们可以以时长作为选项。例如，每天达到8小时以上睡眠则分数为0，6-8小时为1；15分钟内入睡得分为0，15-30分钟入睡得分为1。



以文字作为选项

对于睡眠紊乱、失眠和白天精神状态，可以用文字来作为选项。例如，从不发生表示得分为0，有时发生得分为1。



计算睡眠效率

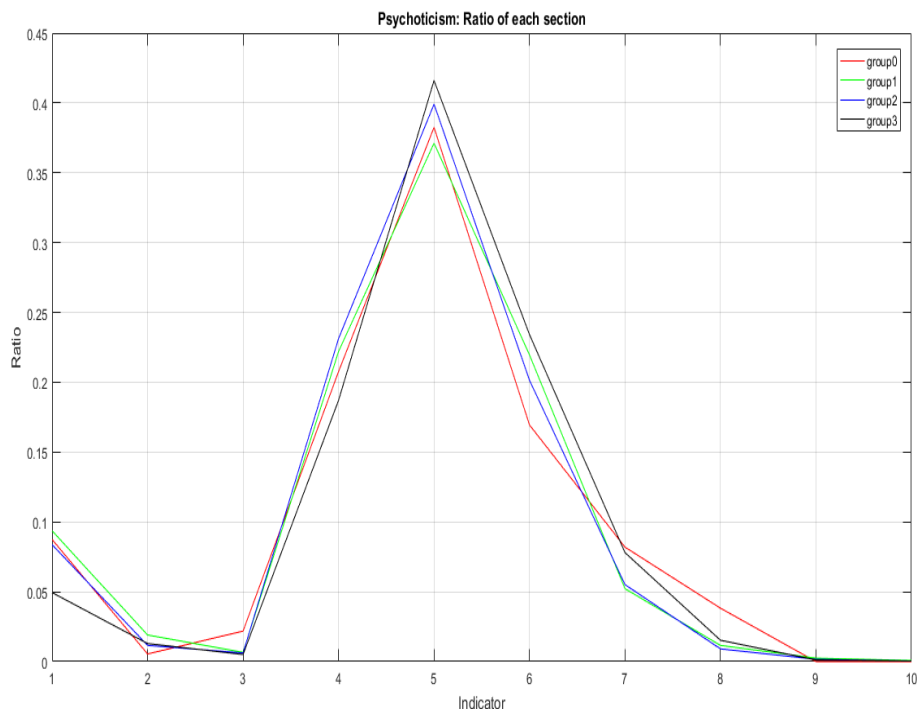
对于睡眠效率，用户当然不懂得如何衡量和计算，所以我们需要通过其他数据计算得到。根据常识，一个人睡眠时长越大，精神应当更好。所以，根据睡眠时长得分与白天精神状态得分之比，推断睡眠效率。

问卷调查表格

Corresponding score Indicator	0	1	2	3	Remarks
Sleep quality	0	1	2	3	
Sleep latency (mins)	<15	15-30	30-60	>60	
Sleep time (hours)	>8	6-8	4-6	<4	
Sleep efficiency	$Eff = \frac{\frac{1}{score(d) + 1}}{\frac{1}{score(t) + 1}} \times 100\%$ $= \frac{score(t) + 1}{score(d) + 1} \times 100\%$ $score(e) = f(e) = \begin{cases} 3, & Eff \leq 0.25 \\ 2, & 0.25 < Eff \leq 0.5 \\ 1, & 0.5 < Eff \leq 0.75 \\ 0, & 0.75 < Eff \leq 1 \end{cases}$				
Sleep disorder	Never	Sometimes	Usually	Always	
Hypnagogue	Never	Sometimes	Usually	Always	
Dysfunction	Active	Need a nap	Sleepy	Bad	

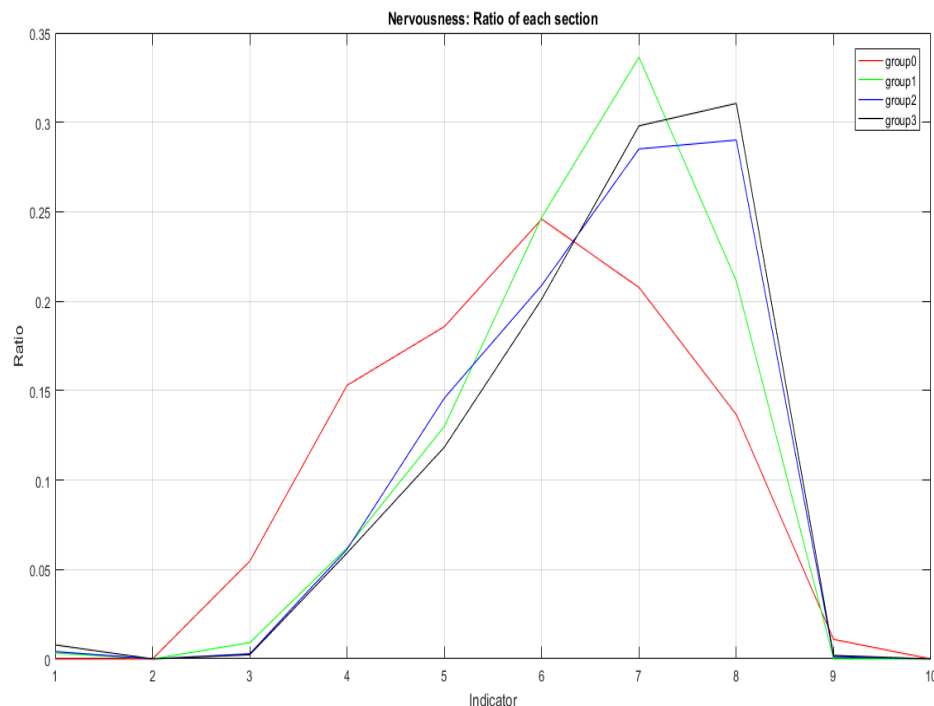
PART
FOUR

问题
求解



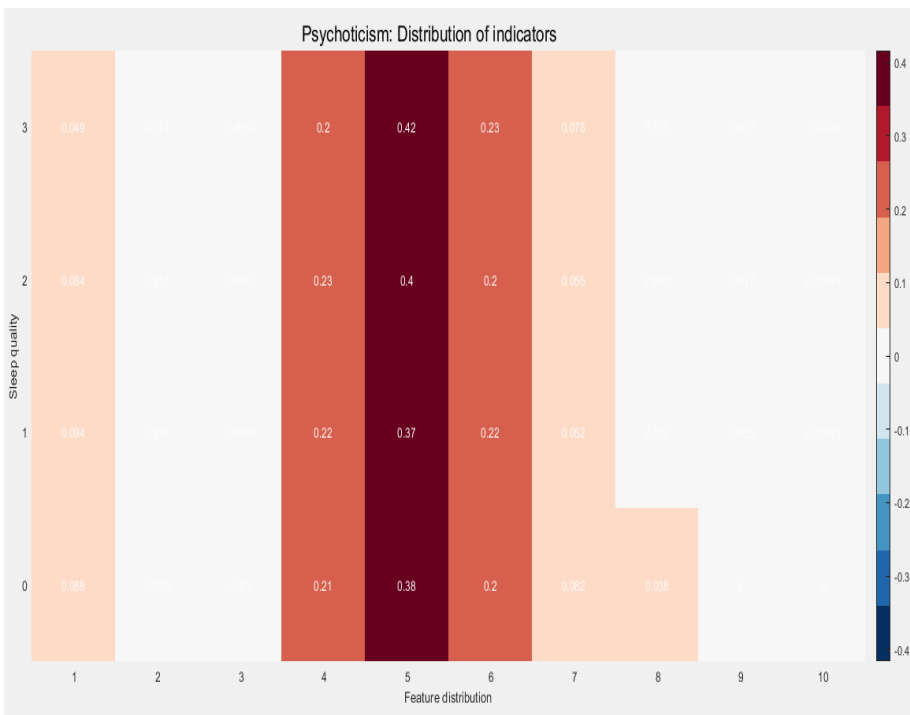
指标Psychoticism下不同睡眠质量人群数量占比

- ✓ 随着指标数据的变化，不同睡眠质量的人数占比接近。说明精神值与睡眠无关。



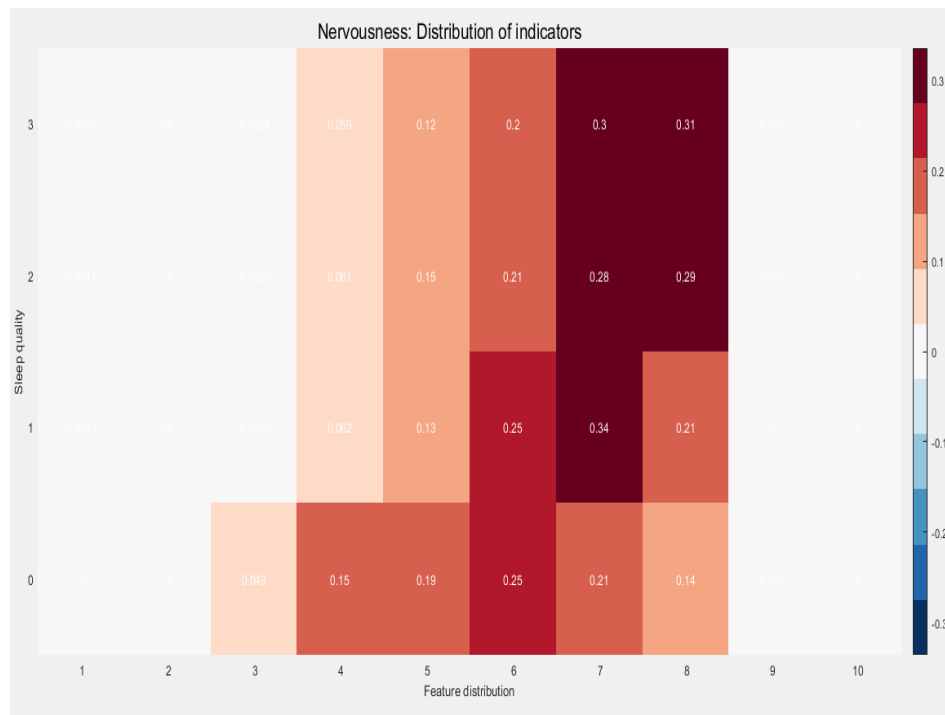
指标Nervousness下不同睡眠质量人群数量占比

- ✓ 指标越低，睡眠质量好的人占比更多。说明焦虑值越低，睡得越好。



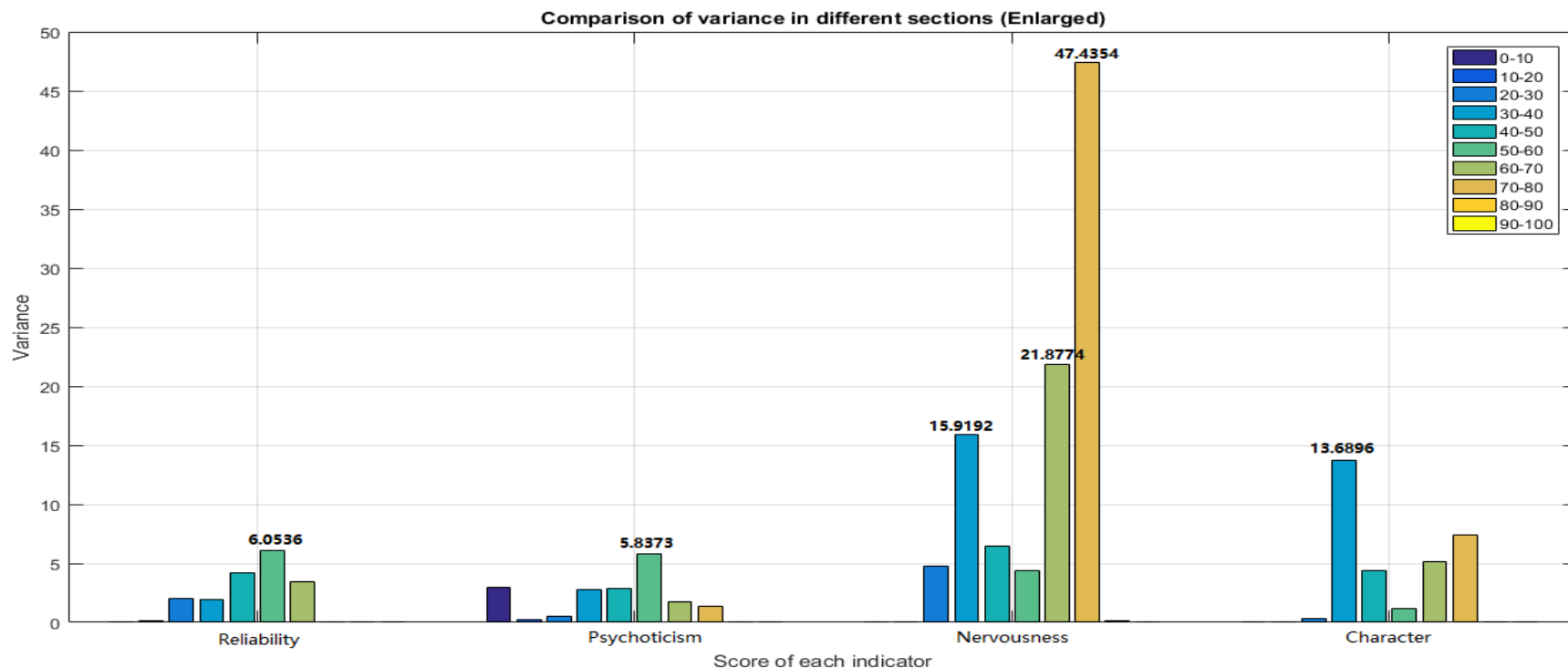
指标Psychoticism下的热度图

- ✓ 深颜色代表数量多。可见随着指标变化，四种睡眠质量的人数占比一直相似。



指标Nervousness下的热度图

- ✓ 指标越高，睡眠质量为2和3的人数更多；指标越低，睡眠质量为0和1的人数更多。



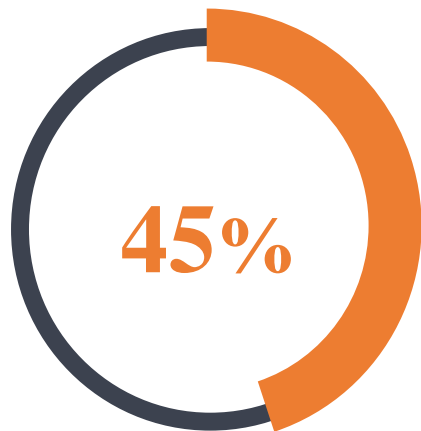
- ✓ 计算四种指标各个得分段下，不同睡眠质量人数占比的方差，得到结果如上图。制定规则为最高方差超过10则代表该指标与睡眠质量有关，否则无关。

Number	Age	Sex	Source	Predicated Diagnosis	Probability
1	28	male	Outpatient	Depression	61.80%
2	37	male	Outpatient	Depression	72.30%
3	45	male	Outpatient	Sleep disorder	92.50%
4	32	female	Outpatient	Sleep disorder	83.10%
5	64	male	Outpatient	Depression	51.20%
6	29	female	Outpatient	Sleep disorder	91.70%
7	42	female	Outpatient	Depression	61.00%
8	36	female	Outpatient	Sleep disorder	54.50%
9	71	female	Outpatient	Mixed anxiety and Depression	94.30%
10	26	female	Outpatient	Sleep disorder	74.30%

PART FIVE

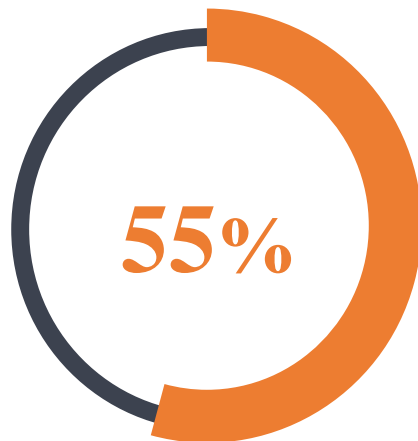
评估 对比

评估对比



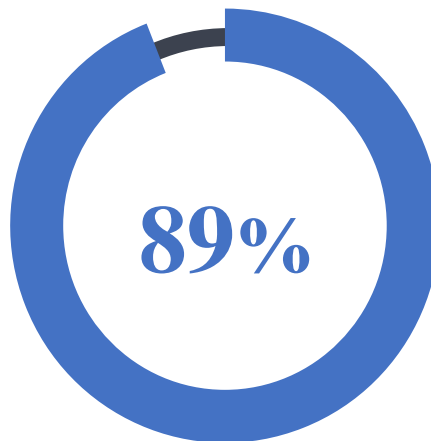
SVM

解决多分类问题时
需要构造多个二类
支持向量机，降低
准确率



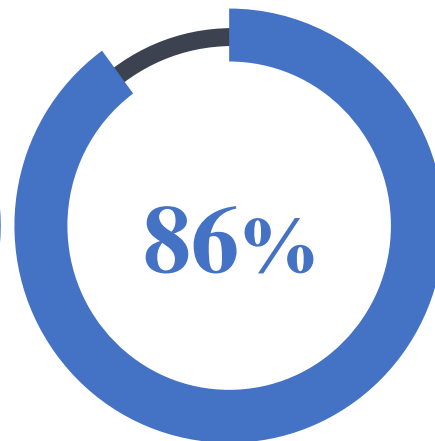
Single Classifier

单分类器易受到
数据偏斜的影响



RDNN1

RDNN第一层
分为四类



RDNN2

RDNN第二层
分为五类

PART SIX

优点
缺点

优缺点分析

优点

结合深度学习知识，模型使用了独创的新型神经网络结构，完美解决了原始数据集中数据偏斜的问题；即使存在数据量较小、特征数量不足等问题，模型仍能给出准确率较高的结果。

模型只能对单个疾病做出预测，且疾病种类的总数目不能过多。此外，对模型诊断患病类型为Others的患者，无法给出精确的结果。

缺点

THANKS