行缺失率20%

A方

B方

缺失率，有什么用

剩下里面80%学习然后对20%进行填补

行缺失率对

行缺失率与与生成规则有什么影响

随机生成强相关的数据

随机生成3列，

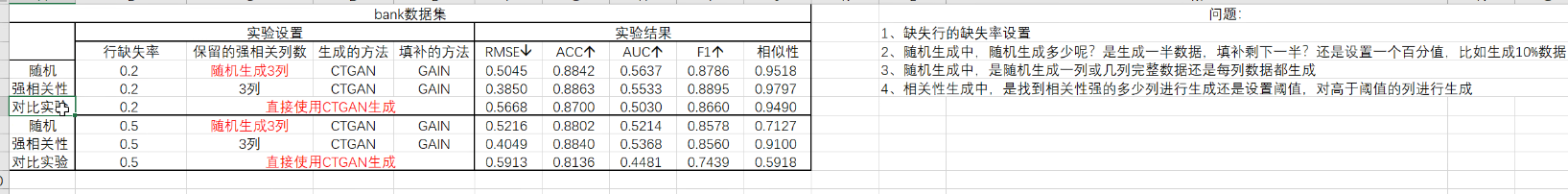
强相关

就用随机函数随机取

对比实验

对缺失的直接生成

CTGAN填补



强相关是怎么比较，不同的是怎么比较。

所有的A，B两方的列与标签列之间的关系。

是不是所有的应用是不受都需要结合

当A方的数据和标签列达到什么关系的程度上可以选中生成方法

A和B的之间的相关性

很多的填补方法

不同的

A方和B方的规则相关还是不相关

相关的怎么来给一个规则

先定一个简单的规则

找几个强相关性的做填补是不一样的。

1. 有没有必要要多方
2. 在确定要多方的情况，这么设计规则

规则的负责程度

半监督有没有可能应用在这里面。

这个规则赋予到

多对一的情况，二分类，多分类

给它一个分类的表示。

细化现有实验

联邦填补的方法

剔除掉null

定了实验设计，就开始跑对照组。

首先标签那个要改

两个要设计实验

阶段性实验是什么

规则：

先做强相关

讨论想法O不OK

基本的填补模型的验证，把那几个模型用熟