学生满意度预测模型

* 1. 学生满意度预测分析流程

LighGBM方法在教育行业学生学习满意度预测问题的应用遵循数据挖掘技术的一般工作流程，包括如下五个部分:(1)确定目的若想建立一个高效的问题处理模型，首先需要了解即将处理的业务对象以及要想达到的目的，据此才能选取合适的客户资料进行特性挖掘。本文工作的目的在于预测各个学生的满意度，从而对学生的情况能够及时做出调整。 (2)数据理解该阶段的主要目标是对已有的相关资料进行理解和熟悉，并将这些资料进行具体的归类分析，为下一阶段的数据准备打下基础。(3)数据准备该阶段的主要目标是将数据理解中归纳的资料进行进一步的处理，首先从全部的资料中抽取出对研究学习满意度当前现状的有利的信息，然后进行数据清洗、分数取值的转换。通过这一处理过程需抽取出直接可用于建立LightGBM模型的分数数据。并确定各个影响因素对应的取值范围.。

(4)建立模型(5)模型评价建立学习度满意度预测模型之后，需要验证模型的效果。具体方法如下:用之前划分的测试集作为输入来测试模型，预测模型会根据每个学生的当前状况预测出一个学习满意度分数。我们将学习满意度分数划分为三个层次：大于等于12且小于等于15为很满意，大于等于7且小于等于11为满意，小于7为不满意。如果实际满意度分数和预测满意度分数在同一个区间，则我们将其鉴定为合格，反之，则不合格。

经过一系列步骤之后，学习满意度预测模型就建立完成了。下面将从上述五部分详细介绍模型的具体建立。

* 1. 学生满意度预测分析的目的

我们可以根据学生当前的外部环境和内部环境，来智能预测出学生当前的学习满意情况，从而可以对学习满意度较低的学生及时地进行帮助。

* 1. 学生满意度预测分析的数据理解

数据理解的目的在于熟悉影响学生满意度的因素，并对影响因素进行归纳，总结，分类，数据理解过程是数据挖掘流程的重要部分，只有对数据充分熟悉的基础之上才能够进行数据挖掘的操作。根据之前的分析，下面对学生满意度影响的所以因素可以从内部因素和外部因素两个角度来看，共有七大因素，分别是

（1）

（2）

（3）

（4）

（5）

（6）

（7）

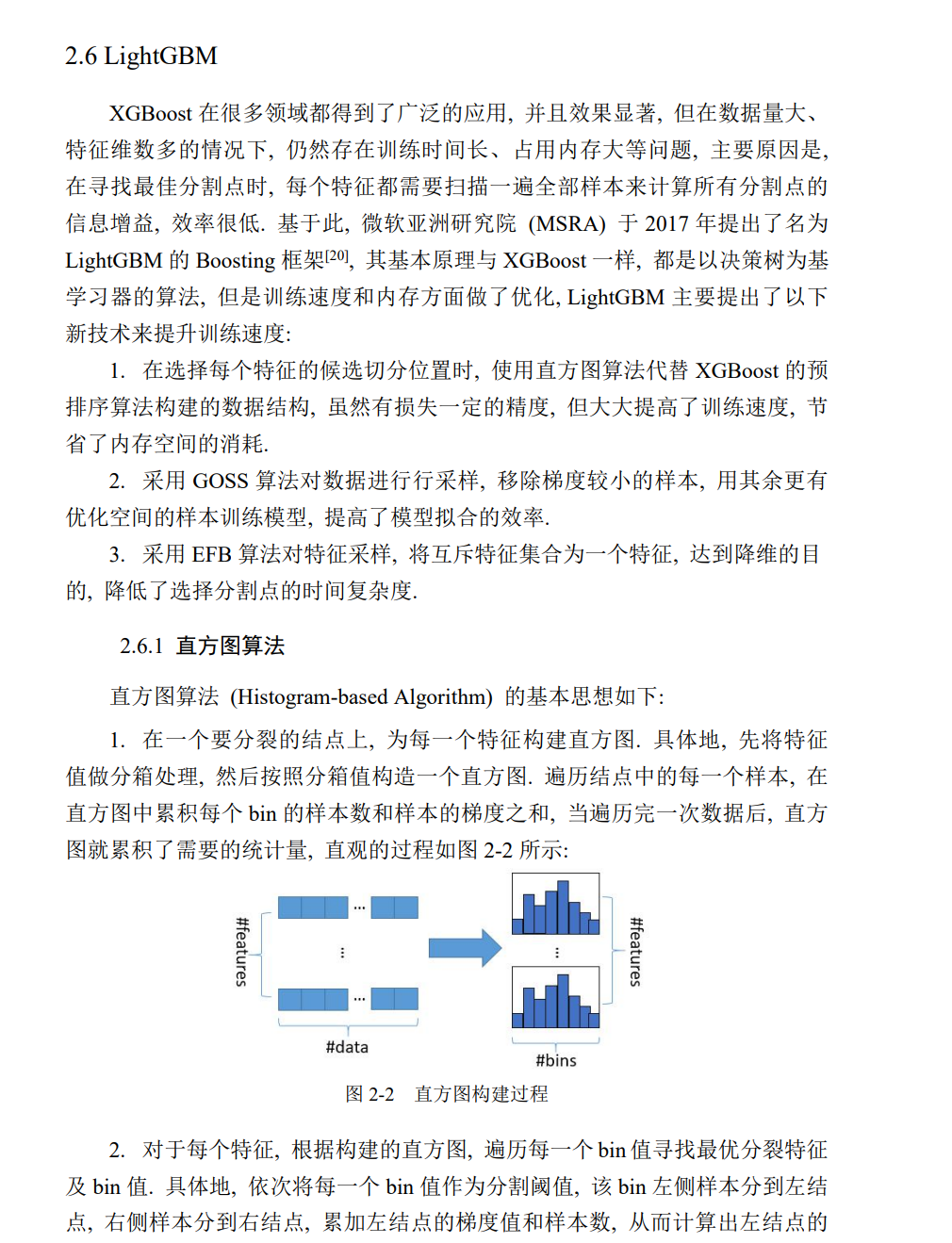
（每个部分进行介绍）

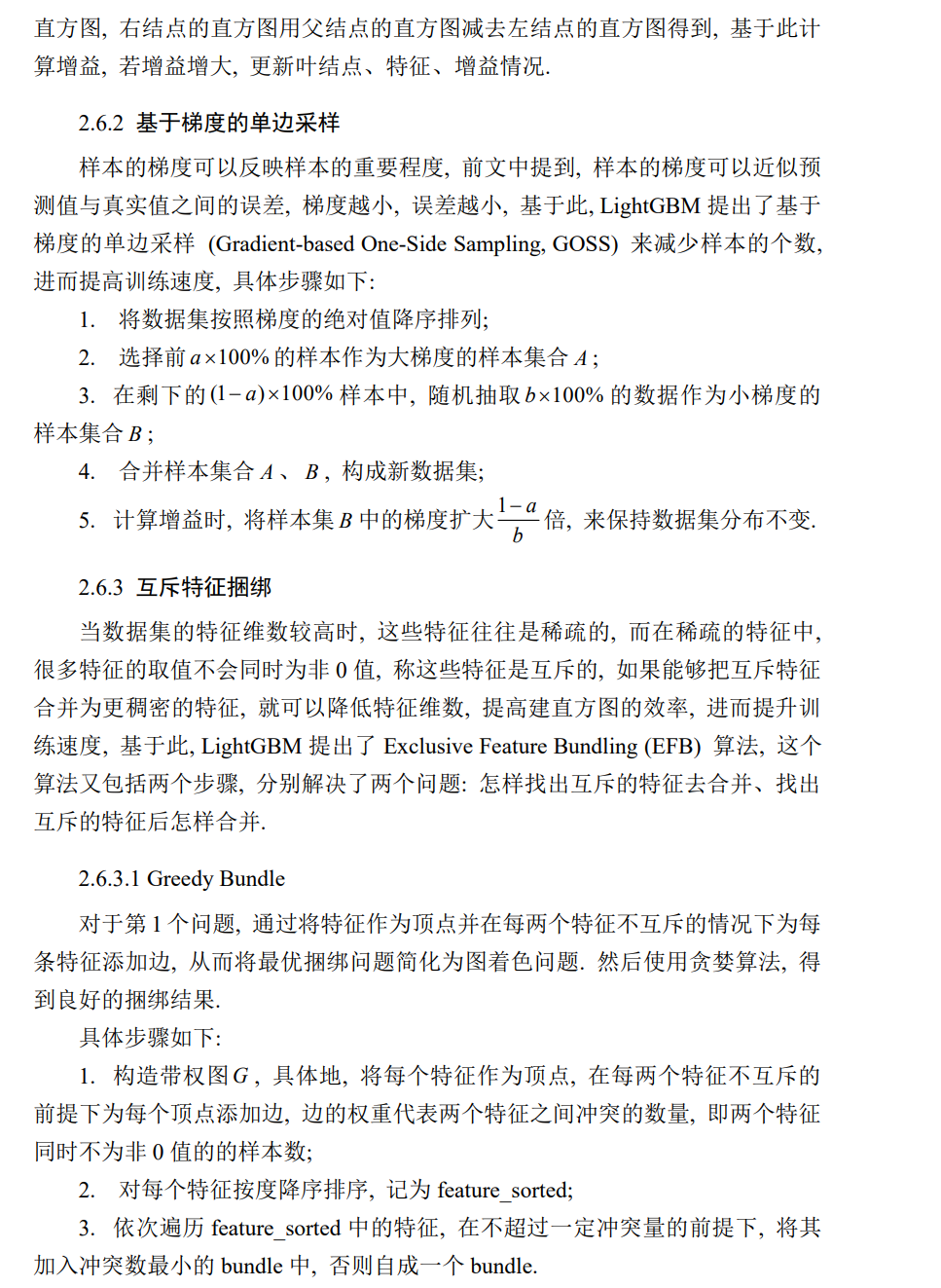
* 1. 学生满意度预测分析的数据准备

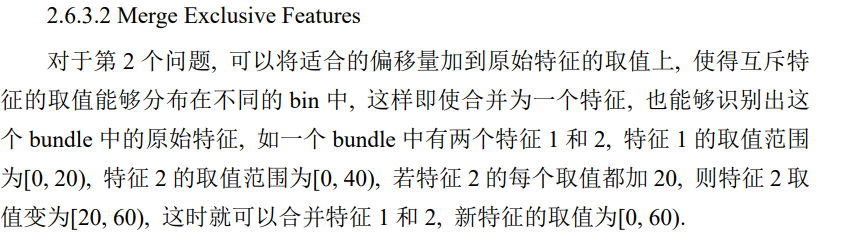
考虑到在满意度影响因素中有些数据是分数越高，其实表示越不满意，例如老师无法做到及时回复电子邮件或帖子，反馈学生关注的问题。这条影响因素，所以我们首先对与这类影响因素的分数取反加上六分得到一个统一的标准，除此之外我们对每个因素下面的所有细粒度的打分进行求和，然后除以细粒度的个数，得到每个因素的均值（相当于是做了一个归一化）。

* 1. 学生满意度预测模型

1.5.1模型介绍（做一个小小的参考，网上专业的详细论文介绍很多）



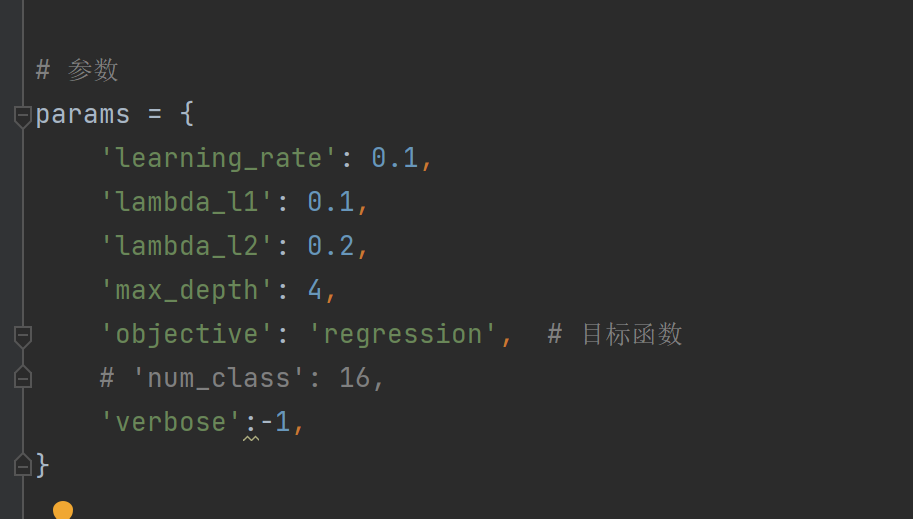




1.5.2 数据集划分

我们将内部影响因素和外部影响因素的所有数据作为一个统一的集合，并按照8:2的比例划分训练集和测试集，其中8的部分我们用来训练我们的模型，其中2的部分我们用来测试我们的模型

1.5.3 模型超参数设置（暂时是这样后面会再修改）



1.5.4跑模型过程

按照论文介绍中的那样放进我们的数据讲一遍流程，然后画一些关于树的图进行辅助分析。

* 1. 模型结果验证

1.6.1验证标准

（可以是1.1中的说的那样划分为三个部分，也可以是预测出来的满意度分数在实际满意度分数的正负2之内就都合格，反之不合格）

我们预测出来的结果如下实际的结果如下：

（这里可以放两张结果表）

可以看出我们模型的合格率是XXX多少，还是比较好的，可以看出啥啥现实意义。