**题目**：银行对中小微企业的信贷决策分析

**摘要：**背景+各个问题的求解思路阐述，重点模型算法用黑色字迹加粗

1. **问题重述**
   1. **问题背景**
   2. **问题提出**

总的来说，第一块十分短小，占用篇幅仅为500字左右，原因为该赛题较简单，问题询问量少，导致问题重述内容少。

1. **问题分析**

**2.1 问题一分析**

基于要求，确定：主要是对不同信誉等级的企业提供相应的贷款额度、利率与是否放贷来制定策略。首先，从四类指标出发，对数据进行特征提取，共提取出20个相关的特征。然后运用Voting集成学习算法，对提取出的特征进行训练，得到每个企业的违约风险值。之后，利用题目给出的信誉评级，算出每个等级的平均违约风险，以此代表该等级企业的违约风险。最后建立以银行利润最高，银行风险最低，潜在流失客户率最小为目标的多目标非线性规划模型。求解得到结果。

**2.2 问题二分析**

强调本问与第一问的主要差别在于本问的企业无信贷记录。解决方法为运用第一问训练出的的 Voting集成学习算法，求出302家企业的违约风险，对违约风险大于50%的企业不予放贷。同时，确定企业的信誉评级是一个多分类问题，将企业的违约概率这一特征与之前提取20个特征组成新的21个特征，运用Xgboost模型，用附件1所提取的特征训练模型，然后使用附件2数据对302家企业进行信誉评级，得出分类结果。最后求解多目标非线性规划模型，解出银行信贷总额为1亿元时对不同信誉等级的企业发放的相应贷款额度与利率。

**2.3 问题三分析**

强调选择的特殊事件为新冠疫情事件，依据疫情对企业的影响程度对302家企业进行聚类分析，配合问题二所做出的信誉评级，建立一个新的综合分类评级。不同类别的企业具有不同的风险乘数Qᵢ。在之前的多目标规划模型中，加入风险乘数Qᵢ，对银行的风险目标进行修改，得到调整后的多目标规划模型，解出模型后得到银行在当期情况下的信贷策略。

1. **模型假设**

1、同一信誉评级下，银行将给予相同信誉评级企业相同的贷款额度与年利率。

2、针对有信贷记录的企业，银行根据以往的信誉评级给贷款。若该企业之前就有违约记录，则今年不予以贷款。

3、针对无信贷记录的企业，银行根据已有信贷记录企业的数据对于无信贷记录企业进行信誉评级之后再考虑是否贷款。

4、银行给予贷款的目标为，所得利息收益最大，承受的风险与客户流失率最小，银行根据此目标对贷款企业发放贷款额度和给予利率优惠。

5、题目附件中的数据真实有效，全面无遗漏。

6、不论利率多高附件内企业都不会放弃贷款。意思是强调企业是请求贷款者，决定权在银行手中。

1. **符号说明**
2. **问题一建模与求解**

**5.1 数据预处理**

运用 Python 的 pandas 库, 找出缺失值。但发现无缺失值，无处理。

**5.2 数据特征提取**

根据附件一所给出的数据，结合**相关文献**和题目，本文认为银行放贷额度与利润率的选择主要受企业自身实力、发展前景和对上下游企业的影响力这三个方面进行评估。根据附件一所给信息，本文就四个方面总共提取了20个特征，用以衡量银行对企业信誉等级的评判，以此来衡量企业贷款的信贷风险。

**5.2.1 企业实力**

首先，我们可以得出企业在所给日期之内**所得利润总额**。再得到企业在所给日期之内的**增值税Tax**，增值税为负的情况置零。得出在所给日期内的**作废票据比**。

**5.2.2 企业对上下游企业的影响力**

根据所给的开票日期判断各个企业与其上下游企业合作的时间长度由此来判断供应链的强度。强度与粘性越大，即代表企业对于上下游企业的影响力越大。对于这种企业银行可适当提高贷款额度，降低利率。

**5.2.3 企业的发展潜力**

2020年的数据并不为全年数据，在计剔除了2020年的票据数据。分别求出每个企业的利润绝对数的变化An和相对数的变化Rn。当Rn小于0时，则代表企业利润发生了正负转变，这时候需要观察An。当An大于0时，说明企业最近年份利润为正而最早年份利润为负，企业扭亏为盈， 即ToPro为1，反之为0； 当An小于0时，说明企业最近年份利润为负而最早年份利润正，企业由盈转亏， 即ToLoss为1，反之为0。针对扭亏为盈的企业，银行可适当提高贷款额度与降低利率，针对扭赢为亏的企业，银行可适当降低贷款额度与提高利率。

**5.2.4 企业的抗风险能力**

将所有企业划分为四种情况内。当公司为独立公司时，Independent为1, 当公司为个体经营, Individual为1, 当公司为子公司时,Controlled为1, 当公司为下属部门, Under为1。

**5.3 信贷风险评分模型**

**5.3.1模型选择**

本文主要参考传统信用评分方法中的多元非线性回归模型。特别的，强调了利用参阅参考文献，了解综合其它论文方法，同时本文还采用集成多个机器学习算法的方式，进行信贷风险度量。

**5.3.2 数据选择与处理**

模型将使用前文从附件一中提取出20个特征，共123个样本进行训练。使用Python将数据随机切分为训练集和测试集，并以原数据为标准，对训练集和测试集数据进行标准化，将数据处理至0到1之间。

**5.3.3 模型建立**

投票(voting)是在分类算法中广泛运用的集成学习算法之一。本文所选择的 5个基分类器分别为Logistic 回归, Adaboost, GBDT, SVM和随机森林。然后文章，简要介绍各基分类器的原理与各算法的工作方式，结合本文具体实例给出公式，是模型解释的核心。

图1 Voting模型图解（简单框架流程图）

**5.3.4 模型求解**

表1 基分类器分类效果

基于分类效果做出评价分析，指出：训练集和测试集效果最接近的为 Adaboost，其余的分类器有程度不一样的过拟合现象。在整体上，Logistic 回归明显略逊一筹，后三个分类器的结果较为接近，主要是三个分类器在训练集上都有良好的表现，而训练集数据占到了整体数据的7成。然而仅从准确率的角度判断分类模型的结果是有失偏颇的，并且本文的目的是，所以本文加入了AUC 值进行参考。AUC 值为ROC 曲线覆盖的面积，其含义可以综合的考虑召回率(recall), 精确率(precision)和准确率(accuracy)多种指标, 一般AUC 值达到0.8以上分类模型可以接受，从这个角度来看，Logistic 回归不符合要求。

因为各个基分类器的分类效果不一样，所以本文选择软投票，依据分类器在整体数据上的综合准确率来判断各基分类器的权重。

表2 软投票结果

可以看出，Soft Voting在分类上的表现十分完美，非常接近完美分类器，说明集成学习算法在此数据集上有着优秀的表现。

图2 Soft Voting的ROC曲线图

Voting模型不仅可以判断出企业是否户违约，并且可以给出企业违约的概率，本文将在后文以此作为对企业的信贷风险的度量。依据信誉评级分类，对不同类别企业的违约风险进行描述性统计分析，表格如下：

表3 不同信誉评级下违约风险的描述性统计分析

**5.4 信贷策略多目标规划模型**

**5.4.1 模型分析**

首先， 由假设1可知银行认为同种信誉评级下的企业的违约风险一定，且在前文中已经求出了每种类别的平均违约风险,故由每类企业的平均违约风险代表该类信誉评级下企业的违约风险。

**5.4.2 模型建立与求解**

三个目标：收益目标（收益最大）；风险目标（风险最小）；潜在损失目标（损失客户量最小）

**查阅资料，我们发现银行有对贷款额度进行分级的实践，不同信用等级的企业能贷的最大贷款和最低利率不同，所以银行应该对信用等级高的企业优先给予贷款，且提供数额较大，利率较低的贷款，以满足银行对收益和风险的追求。**本文因此对不同信誉评级的企业贷款的额度和利率的上下限做出规定，以此来规范贷款行为，具体规则如下：

表4 不同信誉等级下的贷款限制

银行在贷款额度为固定数额的约束下(本问定为5000万)，要使得满足上述三个目标，将上述三个目标转化为单一目标，列式，运用 MATLAB 求解。

结果：对D级企业和有违约经历的企业拒绝提供贷款；剩下的企业中，对A级企业提供利率为5.85%，额度为999905元的贷款；对B级企业提供利率为7.5%，额度为535178元的贷款；对C级企业提供利率为11%, 额度为100030元的贷款。

1. **问题二建模与求解**

**6.1 数据处理与编码**

对附件2进行和附件1一样的数据处理方式，提取出问题一所列的20个特征。

将企业的信誉评级数据进行编码，若信誉评级为A，则Grade=1；若信誉评级为

B, 则Grade=2; 若信誉评级为C, 则Grade=3;若信誉评级为D, 则Grade=4。

**6.2 模型建立**

**6.2.1 违约风险预测**

直接根据问题一中模型，最终得到302个违约风险中，有34家企业的违约风险超过了50%，即认为会违约，银行将不会向他们提供贷款。

**6.2.2 信誉评级鉴定**

本问的企业无信贷记录，所以选择基于XGBoost模型对企业进行信誉评级。

介绍XGBoost模型，写出具体算法步骤：

1、优化目标。2、对L(t)进行二阶泰勒展开和移除常数项操作后可得算式，调用Python的xgboost库，对21个特征(附件2提取出的20个特征加上是否违约)进行训练。XGBoost在附件1上的综合准确率为0.878，在多分类问题中表现较好。

将信誉评级的分类结果和违约风险的分类结果进行比较验证，发现所有的D级企业均为预测会违约企业，除此之外只有两个C级企业预测会违约。两个不同模型训练的结果较为相近，进一步验证了本问分类的精确性。结果被加入附录。

**6.3 信贷策略目标规划**

总结算式，运用 MATLAB求解。对D级企业和有违约经历的企业拒绝提供贷款；剩下的企业中，对A级企业提供利率为7.05%，额度为999717元的贷款；对B级企业提供利率为7.52%，额度为260320元的贷款；对C级企业提供利率为11.01%, 额度为100048元的贷款。

1. **问题三建模与求解**

**7.1 风险及其传导机制**

经过合理性分析之后，本文根据企业所受系统性风险(行业风险)和非系统性风险(自身表现)对企业的进行聚类分析，同一类别的企业拥有相同的风险乘数，代表该类企业的违约风险较疫情未发生时增大的多少。银行将风险乘数加入信贷策略的目标规划模型，重新制定信贷策略。

**7.2 数据处理和特征提取**

**7.2.1 企业状况**

用2020年1、2月份的销项金额减去进项金额即为盈利情况。

**7.2.2 利润同比增长率**

通过计算2020年1，2月份的企业利润与2019年1，2月份的企业利润，可以得出企业2020年1, 2月份的利润同比增长率。

**7.2.3 废票占比变动率**

废票占比变动率即为2020年1, 2月份的企业废票与总票据比例较2019年1，2月份企业废票与总票据比例的变动率。

**7.2.4 交易企业数量变动率**

公司在受疫情影响情况下，与公司交易企业数量的变动程度。

**7.2.5 行业风险**

**通过国家对于行业的标准化分类方法，本文通过观察企业名称将302家企业一共分为了13个行业。针对其中12个行业，我们通过查询同花顺上的行业股票数据，得到了12个行业股票指数2020年2个月的增长率，用以衡量行业打击程度大小。额外的一个行业中的企业均为为个体经营企业，根据第一财经发布的调研报告，个体经营企业预期降幅约为1/3，本文以此数据作为对行业风险的度量。**

**7.3 聚类模型**

本文将上面计算的5种特征，作为系统聚类的依据。系统聚类首先将每一个样本都分为一类，然后不断计算子类与子类之间的距离，逐渐将所有的子类合并为一个大类。系统聚类算法的流程如下图所示。

图3 系统聚类流程图

将数据代入SPSS，进行系统聚类，为了确定聚类类别K，画出聚类系数的折线图。

图4 聚类系数折线图

根据聚合系数折线图，可以看出，令K=3，聚类效果最好。得出聚类结果后，对聚类的结果分别进行描述性统计分析，可得下面三表。

表5类别1 企业特征描述性统计分析

表6类别2 企业特征描述性统计分析

表7类别3企业特征描述性统计分析

根据以上聚类出的三类行业的表现，本文设置其对应的风险乘数

**7.4 信贷策略**

使用MATLAB解出答案

表8 银行信贷调整策略

1. **模型评价**

**8.1 模型优点**

挖掘处理了题目告诉的信息，考虑全面；使用强分类器，构建 Voting 集成学习算法，AUC 值接近完美模型，得到的违约风险可信度高；多目标规划不仅考虑了常见的风险和收益，也考虑了潜在客户流失所造成的损失；考虑疫情对企业影响时，综合考虑了系统风险和非系统风险的影响。

**8.2 模型缺点**

对于客户流失率选择的时候，一段区间内的贷款年利率用的是一个客户流失率，没有考虑小范围年利率变化时的客户流失率；特征提取过多，有些变量互相有相关性。

**8.3 模型改进方向**

可以使用函数对附件3的数据进行拟合，将离散的客户流失率连续化，多目标规划的结果会更精确；在不对分类效果有较大影响的前提下，可以考虑使用因子分析简化特征数量。

1. 参考文献

略

本文图像不多，值得借鉴的地方有一点就是——参考文献引用的非常多，它利用参考文献解释完善了题目为给到的背景，以及研究建立模型时提取特征的一些原则方法也是借鉴了文献，全文几乎是在各个地方都使用了文献，然后我们也可以在解释问题结果，以及穿插叙述过程中使用参考文献时整个论文更具有严谨性。