2022年南京信息工程大学第十六届数学建模竞赛赛题

（请先阅读“南京信息工程大学校内数学建模竞赛论文格式规范说明”）

**B题 蘑菇有毒无毒的分类与识别**

可食用无毒蘑菇因其营养丰富、味道鲜美，是高蛋白、低脂肪，富含人体必需氨基酸、矿物质、维生素和多糖等营养成分的健康食品，深受广大人民的欢迎。然而，自然界中蘑菇种类繁多，尤其很多有毒蘑菇与可食用无毒蘑菇看起来非常相似，导致每年都有误食毒蘑菇中毒的事件发生，也是我国食物中毒事件中导致死亡的主要因素之一。因此，蘑菇有毒无毒的分类与有效识别受到了国内外科研工作者的广泛关注。

附件secondary\_data.csv给出了61069个蘑菇数据，每个蘑菇都被明确分为无毒或有毒，同时给出了每个蘑菇菌盖、菌褶、菌柄、菌托等的颜色、形状，以及栖息地和季节等各类指标数据（有关数据与指标信息详见secondary\_data\_meta.txt）。要求你的研究团队根据附件中的蘑菇数据，建立数学模型研究下列问题：

1、针对附件secondary\_data.csv中蘑菇数据，进行数据预处理，如数值规约、属性规约、数据变换、缺失值处理等，并分析蘑菇各指标的基本特征；

2、结合蘑菇的有毒无毒，分析其与各指标的相关性，讨论有毒无毒蘑菇在哪些指标上存在显著差异；

3、结合上述分析，构建蘑菇有毒无毒的分类与识别模型，对相关因素的重要性进行排序，确定影响蘑菇毒性识别的主要因素；

4、利用secondary\_data.csv数据，从影响因素选择、模型方法选择等角度，评估上述所构建判别分类模型在蘑菇有毒无毒方面的识别效果，并开展模型灵敏度和可靠性分析；

5、针对有毒蘑菇进一步进行分类，并分析对应类别下蘑菇的特征；依据所研究成果，给出有毒无毒蘑菇识别的参考建议。

注：除了附件提供数据，也可以自己进一步采集增补相关数据完成赛题。

参考文献：Dennis Wagner, Dr. G. Hattab, 'Mushroom data creation, curation, and simulation to support classification tasks' in Scientific Reports on 14.04.2021