

# 2020 泰迪杯数据挖掘竞赛 C 题评阅要点

此要点仅评阅参考，具体由评委根据实际情况进行评阅

---

本问题来源于实际的问题，基于自然语言处理技术的智慧政务系统已经是社会治理创新发展的新趋势。需要学生针对实际文本数据，应用文本挖掘方法进行建模研究，有针对地解决问题，对提升政府的管理水平和施政效率具有极大的推动作用。

## 问题 1：群众留言分类

本问题要求对群众留言分类，这是一个针对文本的非均衡多分类问题。需要对文本数据进行处理，提取相关特征，对留言进行分类，并给出明确的分类结果。

如果能够依据数据的特点，针对不同的方法进行比较分析, 值得鼓励。

## 问题 2：热点问题挖掘

本问题要求提取热点问题。需要有明确的问题热度的度量，一个好的度量不仅要考虑留言数目、赞成数目、反对数目等数字指标，而且要考虑该问题对社会和政府部门工作的影响。

## 问题 3：答复意见的评价

本问题要求对政府部门的答复意见进行评价，是一个开放性的问题。可以从答复的相关性、完整性、可解释性、即时性等角度进行单独或者综合评价，需要言之有理。

如果能够对问题进行分类，按照不同类别的问题分别进行评价，值得鼓励。

本次评阅采用 10 分制进行评分，各问题的展开应该在数据分析的基础上进行。