**《数据科学与数据分析》**

**项目初步提案**

**（一）组员信息**

|  |  |  |
| --- | --- | --- |
| * **组长** | 罗一逖 | 学号：231820309 |
| * **组员** | 吴林洁 | 学号：231820307 |
|  | 邰培峻 | 学号：231820013 |
|  | 寇琪 | 学号：221300047 |

**（二）项目初步提案**

|  |  |
| --- | --- |
| **项目名称** | **《死生之大——基于死刑数据集的探索性研究》** |
| **研究问题**  **（100字以内）** | 聚焦死刑制度的内在属性和外在影响,探索死刑的意义和价值。具体而言，通过探索死刑的社会歧视性分析死刑制度的内在建制；通过对遗言和群众态度的分析探索死刑外在的个体影响和群体价值。 |
| **可能的数据来源以及数据解决方案**  **（500字以内）** | **1.德州死刑犯遗言数据**  来源TDCJ的公开数据（1982年至今），共覆盖593例死刑执行案例，包括死刑犯年龄、性别、种族、地区、教育水平等信息,前科、罪行、受害者情况等犯罪具体信息，及犯人的遗言数据。  通过人工抽取遗言特征词，分析语义将其归类为不同的悔过程度,运用决策树分类算法将遗言分类，并赋予不同的权值，从而计算死刑犯的悔过度，判断死刑执行对罪犯产生的效用。  **2.twitter话题数据**  来源github项目数据集，包括twitter上所有有关死刑话题共计8000条，数据完整包含了话题的标题、话题的具体文本内容、用户姓名、发布日期。  本研究将通过SentenceTransformers模型对该话题文本进行向量化，以对话题进行语义发掘及聚类分析，并通过LDA模型进行用户建模，发掘用户对于死刑的情感态度特征，探究死刑的群体影响。  **3.美国建国以来的死刑记录**  来源于DPIC，包括自美国建国以来至2025 年1月1日的所有死刑记录，共计24408条；包含了犯人姓名、别名、性别、种族、所处州和县、死刑判决年份等。  **4.美国建国以来的死刑冤案记录数据**  来源于DPIC的Innocence数据库，共200条记录，包含无罪释放的犯人的姓名、州、种族、免罪程序、原因等数据。  3、4数据集会综合用于统计分析探究种族、性别、地区、政治等与判罚的关系。 |
| **预期目标**  **（300字以内）** | 我们预期从内外探究死刑制度的意义。   1. 内在属性：   我们预期从美国建国以来死刑执行记录以及冤假错案记录探索死刑制度本身是否具有歧视性，结合美国历史上重大的死刑制度改革（如1812年、1972年宪法修正案）对死刑执行记录的影响来探索死刑制度本身的沿革意义。   1. 外在影响：   我们预期分析死刑执行产生的个体效果——死刑犯的悔过程度，探究是否具有威慑和教育意义；探索死刑执行的群体价值——民众对于死刑的态度和观点。  本研究期望法官及法律制定者们能够通过该实证分析切实了解死刑制度的意义，以及死刑本身的社会歧视性，以全新的视角去进行法律判决及政治决策，推动死刑制度的审慎适用与系统性改革。做到审慎对待死刑，捍卫正义理想，尊重生命价值。 |