### Situational Analysis

This section provides a detailed assessment of the current state of the project, including a review of available resources, constraints, and a preliminary cost-benefit analysis. This analysis is crucial for determining the feasibility and scope of the project.

#### 1.3.1. Resources Available

The following resources have been identified for the project:

* **Data:** The primary data source is **ZambiaLII**, which provides access to thousands of judicial judgments. The current collection consists of 8,871 judgments stored as large scanned PDFs. A major constraint is that the documents are **image-based** rather than text-based, necessitating the use of **Optical Character Recognition (OCR)** tools (e.g., Tesseract or cloud-based OCR services) for text extraction. A significant challenge is the lack of pre-existing topic labels, meaning all labels must be manually created or inferred from case metadata, such as titles and citations. The open-access nature of ZambiaLII eliminates any financial or legal barriers to data acquisition.
* **Personnel:** The project team comprises five computer science students with a background in data mining. A key constraint is the **absence of a legal domain expert** with specific knowledge of Zambian law. This introduces a **significant risk**: the potential for misinterpretation and mislabeling of legal terminology. To mitigate this risk, the team will rely on publicly available case abstracts and other metadata for cross-checking, but this assumption must be explicitly noted in the project plan.
* **Technology:** The team has access to standard computational resources, including personal laptops and **free cloud-based notebooks like Google Colab**. The project will exclusively use **open-source software**, including Python and various NLP libraries (e.g., NLTK, spaCy, scikit-learn). This approach eliminates licensing costs, but it also means the project is limited by the free GPU resources provided by Colab. Given the moderate size of the dataset (a few thousand documents), these resources are considered sufficient.
* **Project Requirements & Deliverables:** The project must meet a set of specific academic requirements. The primary deliverables include a fully documented codebase in a Jupyter notebook format and a formal, written report. The project's scope must be achievable within the academic semester. A critical requirement is to prepare a dataset of at least a few hundred to a thousand judgments for model training. The data must be representative, encompassing judgments from different courts and time periods to avoid systemic biases.

#### 1.3.2. Cost-Benefit Analysis

A preliminary cost-benefit analysis indicates that the project is a worthwhile endeavor, with benefits outweighing the costs.

* **Costs:**
  + **Student Effort and Time:** This is the primary cost, encompassing the labor-intensive tasks of OCR processing, manual data labeling, and model development.
  + **Financial Costs:** These are negligible. The use of open-source tools and free data sources ensures no direct monetary expenditure. While there is a possibility of incurring costs for paid cloud services, the project plan is designed to operate within the free tiers of available resources.
* **Benefits:**
  + **Educational Value:** The core benefit for the project team is the practical, hands-on experience gained in applying data mining and NLP methodologies to a real-world problem.
  + **Prototype Development:** The project aims to produce a functional prototype of an automated legal topic classifier.
  + **Long-Term Impact:** While not a direct objective of this academic project, a successful model could serve as a foundation for a larger-scale solution that significantly improves legal research efficiency by saving time for legal professionals, thereby justifying a more substantial future investment.