Automated Tagging and Categorization of Incident Learning Narratives in Clinical Settings

Author: [Your Name]

Affiliation: [Your Institution]

# Abstract

This manuscript presents an automated system designed to tag and categorize narratives in incident learning entries within clinical settings. Utilizing Natural Language Processing (NLP) techniques, the system identifies recurring themes, issues, and key elements in the narratives. This automated tagging facilitates subsequent reviews and analyses, offering a scalable and efficient method for managing critical incident reports.

# Introduction

Incident learning systems are crucial in clinical settings for identifying potential areas of improvement, ensuring patient safety, and facilitating continuous learning. These systems often include textual narratives that provide context and detail about each incident. However, the manual review and categorization of these narratives can be time-consuming and subject to human error. This manuscript presents an automated system for tagging and categorizing these narratives, aiming to improve the efficiency and accuracy of incident learning systems.

# Methods

## Data Source

The data for this study is sourced from an Excel spreadsheet containing multiple columns. The column labeled '105.Narrative' serves as the primary data source and includes textual summaries of various incidents.

## Data Preprocessing

The initial data preprocessing involves text cleaning and preparation. This step includes converting all text to lowercase and filtering out non-alphabetic tokens. The preprocessing is essential for the effectiveness of the subsequent feature extraction and tagging mechanisms.

## Feature Extraction

The feature extraction phase involves identifying potential tags that could serve as descriptors for the narratives. These tags include single words, predefined multi-word phrases, and n-grams. The single-word tags are the top 50 most frequently occurring words across all narratives. Predefined multi-word tags like 'Dosimetric Errors,' 'Patient Delayed,' and 'Treatment Planning' are also included based on their contextual relevance. Additionally, frequent n-grams (bi-grams, tri-grams, and four-grams) are extracted to capture sequences of words that appear together frequently.

## Tagging Mechanism

The tagging mechanism utilizes the Term Frequency-Inverse Document Frequency (TF-IDF) technique to convert the narratives into a mathematical representation. This representation is a sparse matrix where each row corresponds to a narrative and each column to a potential tag. The TF-IDF scores are then used to select the top five most relevant tags for each narrative. These tags serve to summarize and categorize each entry effectively.

# Results

The final output is an Excel spreadsheet that includes the original dataset along with a new column containing the top five most relevant tags for each narrative. The output also includes additional sheets listing all the unique tags, frequent bi-grams, tri-grams, and four-grams.

# Discussion

The automated tagging system successfully identifies and categorizes themes and issues in the incident learning narratives. By employing advanced NLP techniques, the system offers a scalable and efficient method for managing critical incident reports in clinical settings. Future work could involve refining the tagging mechanisms, incorporating additional contextual information, and extending the system for real-time analysis.

# Conclusion

This manuscript presents a robust automated system for tagging and categorizing incident learning narratives in clinical settings. The system's utility lies in its ability to summarize, categorize, and thereby facilitate the analysis of these critical reports. By automating this process, the system significantly enhances the efficiency and effectiveness of incident learning systems.