Executive Summary

Unsupervised Wisdom: LLM reveals a direct connection bewteen solitary and falling severity

Email
dalitros@gmail.com

Phone
+972-52-302-0022

Dalit Roth Rosenberg

1. Introduction

Falls among older individuals are a significant problem, causing severe injuries and high medical expenses. To address this, states rely on healthcare providers to identify risk factors and prevent falls. The CDC/NCIPC aims to enhance its understanding of this issue using machine learning, particularly natural language processing (NLP) on medical narratives. This study uses the XXL LLM to extract novel features, focusing on 'was alone' and its correlation with fall severity. Integrating LLMs for feature engineering in narrative text analysis holds potential for rapidly improving our understanding of complex healthcare issues."

2. Problem Statement

The challenge at hand involves utilizing unsupervised machine learning techniques to analyze narratives from emergency department (ED) visits of individuals aged 65 and above who have suffered falls. The main goal is to extract meaningful insights from these medical record narratives to inform policies and interventions aimed at reducing falls among the elderly. The primary objectives include:
a. Investigating how population demographics, such as age, race, and sex, are related to disposition severity. b. Exploring the influence of pre-existing medical conditions on the severity of these cases. c. Identifying correlations between the timing of the incident and its severity.
Achieving these objectives will provide a comprehensive understanding of the relationship between disposition severity and these factors, contributing to more effective strategies for preventing falls among older seniors.

3. Methodology

3.1 Data Preprocessing

Dataset was downloaded and EDA was performed to evaluated the balance of importent features, the distribution of features against a key values and selected the best features that can be work with. (features with too many missing values or repetative feature were not accounted for further research.

3.2 Feature Engineering

In our feature engineering, we aimed to understand factors affecting disposition severity in falls among older seniors. We categorized severity into low (0) and high (1). We added binary features for race (White), age (Under 80), and sex (Man or Woman) to analyze demographic influences. We also created time features to explore temporal patterns and identified top 10 products involved in falls, simplifying location categories with a "home" feature. Additionally, we considered the frequency of "head" in narratives due to its association with severe injuries. These steps enhance our analysis of falls among older seniors.

3.3 Model Selection

LLM model
Analysis by LLM models such as flan-t5-xxl (1-2), that pretrained on a vast amount of text documents, has been used to extract additional latent information from the narrative data (e.g., Activities that were likely performed prior to the accident; health conditions that might have had a direct impact on the fall. etc.) (Fig.1.). This information was added into the primary dataset and was further used for training of supervised/unsupervised ML models to predict some of the main questions such as: can we predict falls (e.g., severity, type of fall injury) among various demographic groups (e.g., by race, sex, age)? and what are the main predictive features that are essential for good predictions.

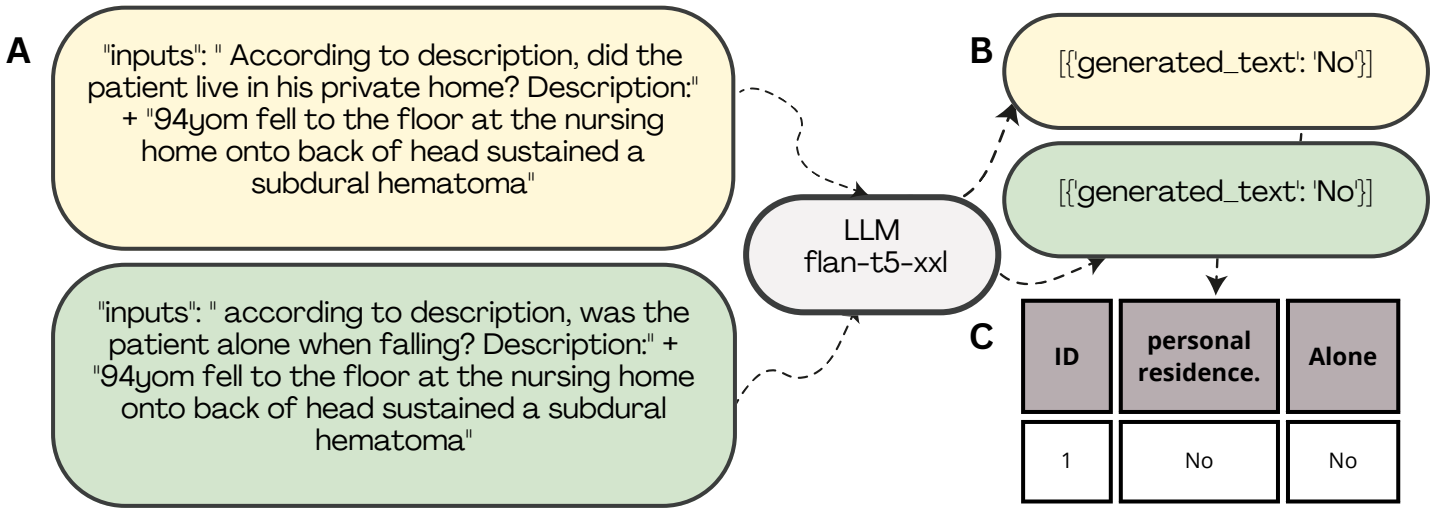

Fig 1. A scheme based on the official model card Inference: *generalization to unseen tasks. The scheme demonstrates several examples of unseen questions (A.). The generated answer (B.) was added as a new feature to the data frame (C.).*

4. Results

4.1 Relationship between Age Groups and Severe Falling Incidents

- We observed differences in the frequency of severe falling incidents between age groups, specifically concerning the presence of being alone.
- Older senior (above 80) who were alone during a fall tended to have a higher prevalence of severe falling incidents compared to younger seniors (below 80).
- Younger seniors (below 80) who were not alone during a fall, especially during holidays, showed a higher prevalence of severe falling incidents (Fig. 2).

4.2 Correlation between Severity of Falling Incidents and Various Factors

- The presence of certain factors such as drug usage, alcohol involvement, being alone, race, and gender was correlated with the severity of falling incidents.
- These factors played different roles in influencing the severity of falling incidents among different age groups of elderly seniors.

4.3 Insights from LLM Analysis

- The LLM analysis revealed hidden information from the narrative text, shedding light on activities performed before the fall and health conditions that might have influenced the fall.
- This additional information was integrated into the primary dataset and used to train supervised and unsupervised machine learning models.

6. Challenges

In our initial analysis, multiple T5 LLM models were tested, with no success until we tried the flan-t5-xxl model. Asking the right questions with this model proved its ability to answer complex queries based on its training.

It's crucial to emphasize that this analysis is preliminary, and we recognize the need for further investigation and fine-tuning to create an appropriate model. Additionally, we intend to rely on the LLM model to eliminate the need for separate text analysis in our future work.

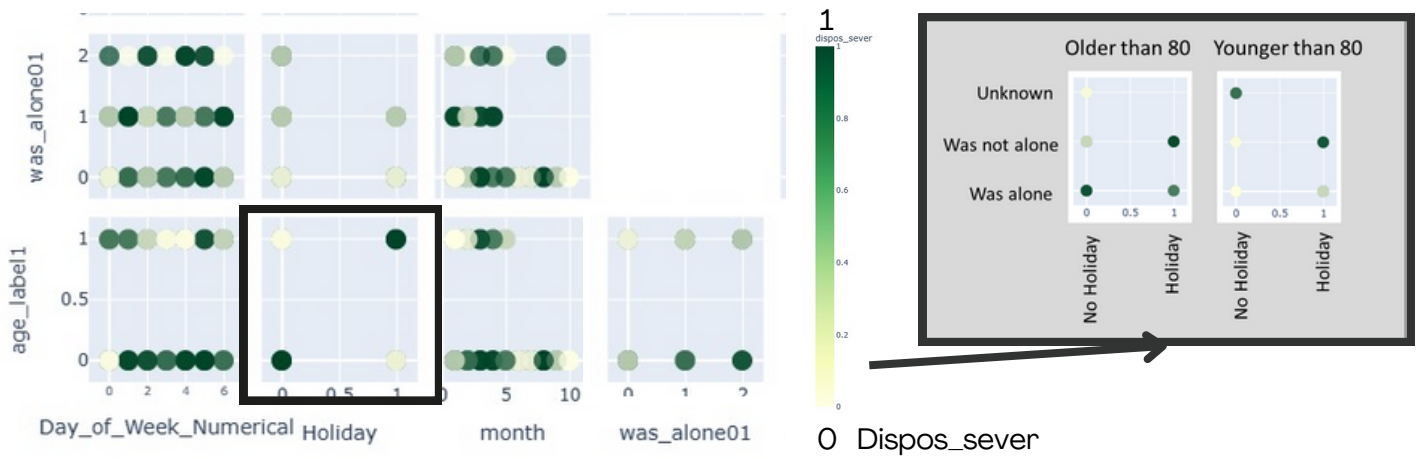

Fig 2. Data Scatter Matrix Plot. *differences in the frequency of severe falling incidents between age groups, specifically concerning the presence of being alone.*

7. Future Recommendations

Future work could involve the development of predictive models using the enriched dataset, incorporating the insights gained from LLM analysis. These models could help in identifying high-risk individuals and tailoring fall prevention strategies to their specific needs.
Further analysis could explore the temporal patterns of falling incidents, taking into account the time of day, month, and year. Additionally, investigating the role of specific medical conditions in falls and their severity could provide valuable insights for healthcare professionals.
Overall, this study provides a foundation for future research on fall prevention among older seniors, leveraging advanced techniques in natural language processing and machine learning to improve the health and well-being of this vulnerable population.

8. Conclusion

The analysis of falls among older seniors, especially those aged 65 and above, is a critical issue due to its significant impact on health and quality of life. By leveraging machine learning techniques, including large language models, we have gained deeper insights into the factors that influence the severity of falling incidents.
Our analysis highlighted the importance of considering different age groups when addressing fall prevention strategies. Solitary seniors are at a higher risk of severe falling incidents, emphasizing the need for social support, safety measures in the home, and addressing underlying medical conditions. Younger seniors are more likely to experience severe falling incidents when not alone and during holidays, indicating the importance of education, home safety, and physical activity.
The use of LLMs to extract hidden information from narrative text has proven valuable in enhancing our understanding of fall-related incidents. This information can inform policies and interventions aimed at reducing falls and improving the well-being of elderly seniors.

CDC Centers for Disease Control and Prevention

DRIVENDATA