Detecting Stress in Clinical Progress Notes

Stress in the healthcare sector has been an ongoing problem for years. A stress monitoring system based on already accessible daily data and natural language processing (NLP) could protect healthcare workers, improve care quality, and reduce costs. This thesis investigates ways of detecting stress in healthcare workers' progress notes and finding patterns in stress predictions.

Introduction

Our work is primarily based on the PsyPN dataset, a non-public collection of progress notes provided by a Swiss psychiatric facility. These notes contain free text and are not annotated for stress. Additionally, notes from the MIMIC-III dataset, created by the Massachusetts Institute of Technology, were used.

Methodology

In the first experiment, a subset of the MIMIC-III dataset, which captures progress notes in a NICU setting, was sampled. After statistical analysis, both datasets were clustered and compared. In a second experiment, previously trained models were used to classify the datasets. These models were created as part of a preliminary project and trained on a dataset annotated for healthcare worker stress. The predictions were then reviewed. For the third experiment, we utilized the temporal information about the notes to create a model of healthcare workers who presumably quit their jobs. For this, a variety of different models were created.

Lastly, we investigated overall stress. Based on the stress classification of the progress notes by our models, we further investigated possible correlations with other data.

Results

Our work confirmed previous findings, indicating that detecting stress specifically among healthcare workers using only non-annotated progress notes is infeasible. Nevertheless, our models demonstrated an ability to detect overall stress levels from these notes. By analyzing the correlation between stress and other data such as local weather, hospital capacity, and COVID-19 data, we further illustrated our model's potential to identify underlying stress patterns within the data.

Possible Applications

The models could be utilized to monitor overall stress levels in healthcare facilities, enabling the identification of periods with high stress and facilitating directed interventions. Additionally, they could be employed for forecasting stress trends and evaluating the effectiveness of implemented countermeasures.
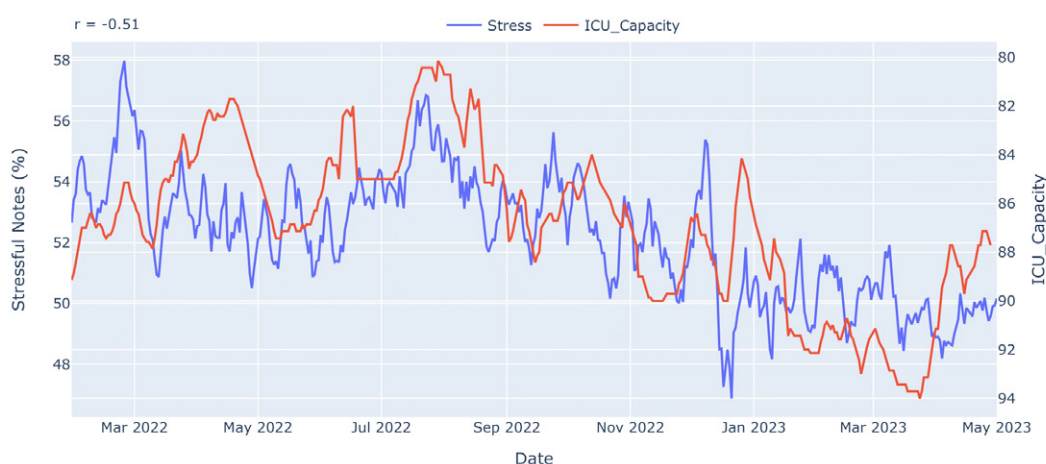
Jasper Benedikt Meinert
Data Engineering

Noah Versace Salvi
Data Engineering

Correlation between Stress and ICU_Capacity in BE (7-day MA)

r = -0.51

Comparison of Daily Stress vs. Number of ICU Beds in Bern Over Time