VAE Insight

VAE-Insight
A Vaccine Adverse Event Prediction System.

<Sikai Cheng, Che-Ting Meng, Priscilla Zhang, Yueming Zhu, Bowen Zuo>

Introduction

Amid the pandemic, Covid vaccines and policies faced significant controversy. Nonetheless, individuals lacked resources to verify the impact of vaccination.

Consequently, some hesitated to vaccinate due to uncertainty regarding its effects on their well-being. However, vaccines often offer benefits such as disease prevention and reduced severity.

We want to help with this problem and help people to prepare themselves for vaccines. If we can have an application that can somehow predict side effects for taking certain type of vaccines for people, we may be able to encourage more people to take vaccines and reduce their risk on health conditions.

Data

The data is from Vaccine Adverse Event Reporting System (VAERS). The data is stored in csv files downloaded. We used about 40000 rows of raw data which contains around 40 features each row.

VAERS
Vaccine Adverse Event Reporting System

Data cleaning:

- Complement missing values with intuition, mean value, KNN, XGBoost and count encoding.
- NLP Processing: Used YAKE! and term frequency to obtain symptoms.
  - Input: Vectors of length 21 (patient & vaccination features)
  - Output: Vectors of length 25 (symptoms)

Complement Method	***Columns***
Intuition	ER_VISIT, HOSPITAL, DISABLE, RECOVD, HOSPDAYS
KNN / XGBoost	VAX_TYPE, VAX_MANU, VAX_DOSE_SERIES, VAX_ROUTE
Mean Value	NUMDAYS, AGE_YRS
Count Encode	LAB_DATA, OTHER_MEDS, CUR_ILL, HISTORY, PRIOR_VAX
No Complement	SEX, FORM_VERS, V_FUNDBY, V_ADMINBY

Models & Experiments

We tried LSTM and CNN models for our problem at first. We then improved from those vanilla models and built a MLP with the structures above. To predict the results, we had two metrics. The first one is mean absolute value evaluation between prediction and true. The second is the exact matches for all 25 dimensionalities.

There isn't an exact competing method out there compare to our application. In general, due to the nature of our problem. The results are mostly for reference. So our prediction have much stronger and more direct meaning for people.

Model:

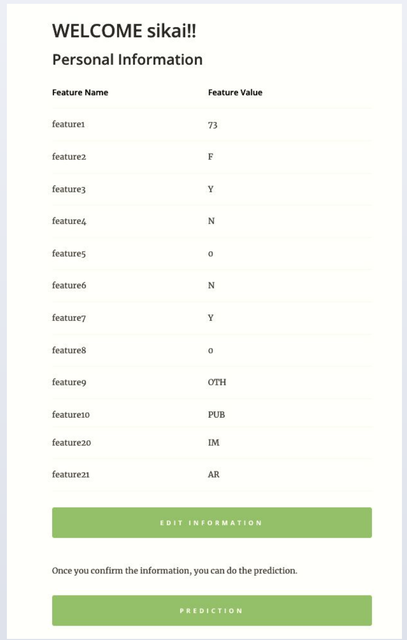
- Multi-Layer Perceptron (MLP)
- Long Short-Term Memory (LSTM)
- CNN (Convolutional Neural Network)

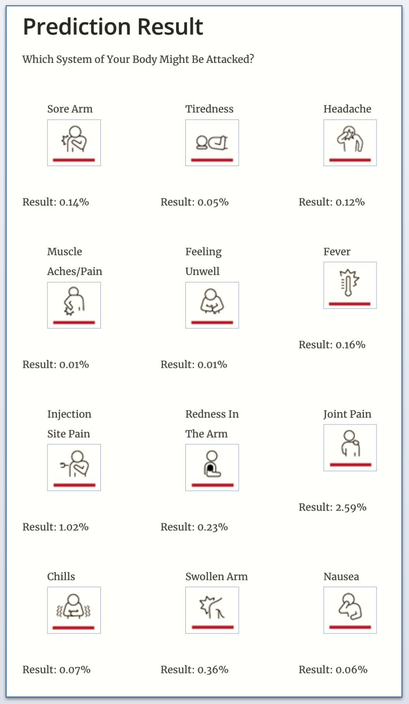
Model	MAE evaluation	Exact match accuracy
LSTM	79.48%	18.57%
CNN	81.58%	17.53%
MLP	**94.39%**	**26.75%**

Layer	Input	Output
Dense	21	128, ReLU
Dense	128	64, ReLU
Dense	64	32, ReLU
Dense	32	25, Sigmoid

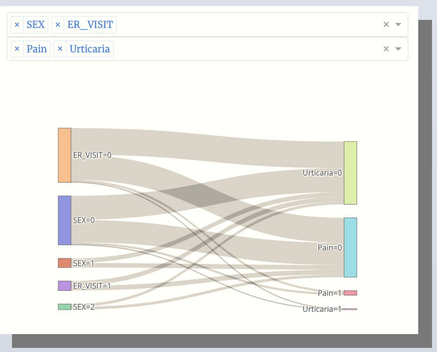
UI Design

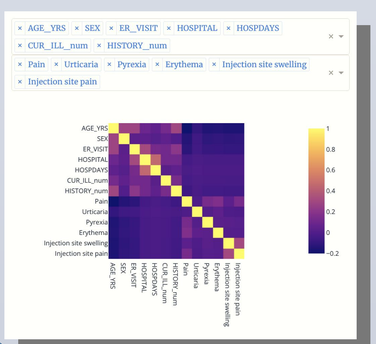
The predictive algorithm generates forecasts pertaining to 12 prevalent adverse events that may arise following vaccine administration.





The Sankey diagram will show the correlation (causality) between input features and output symptoms in the form of flows, and the width of the flows will intuitively reflect the size of the different links.





The heatmap plainly visualizes the cor- relation between the selected variables using a set of color gradients.