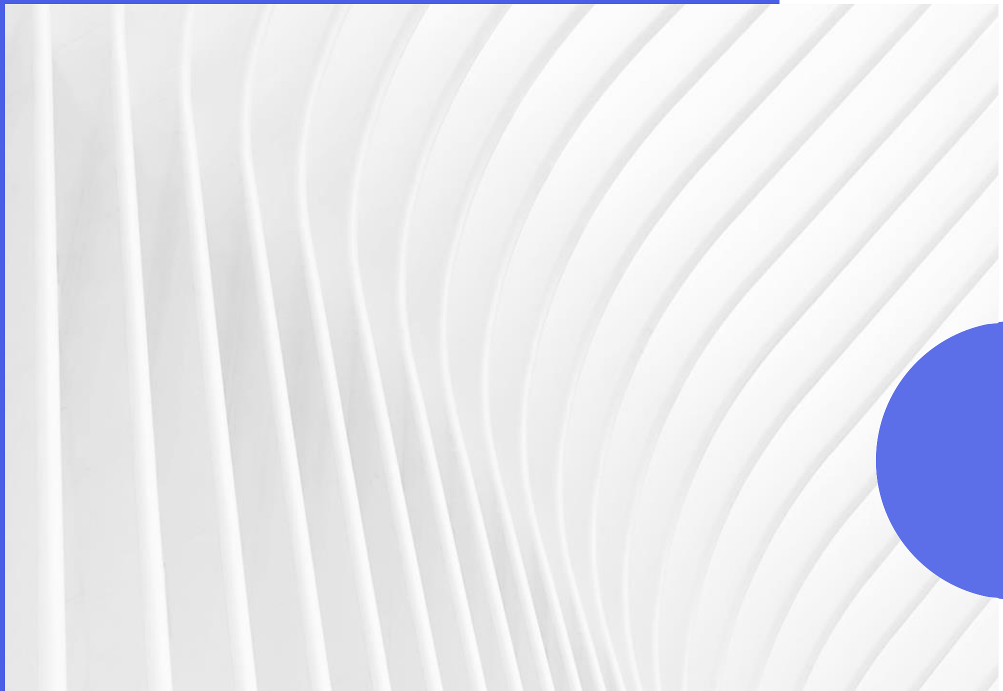
TEAM MEDBOTS

MED-NLP PROJECT

UMBC

MOTIVATION



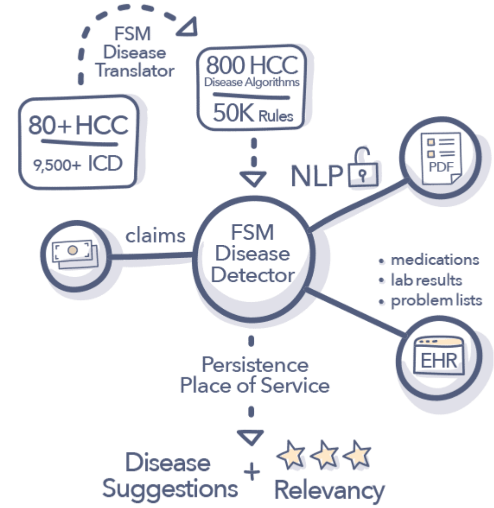
"Many online patient reports are not coded but are recorded in natural-language text that cannot be reliably accessed. Natural language processing (NLP) can solve this problem by extracting and structuring text-based clinical information, making clinical data available for use."[1]

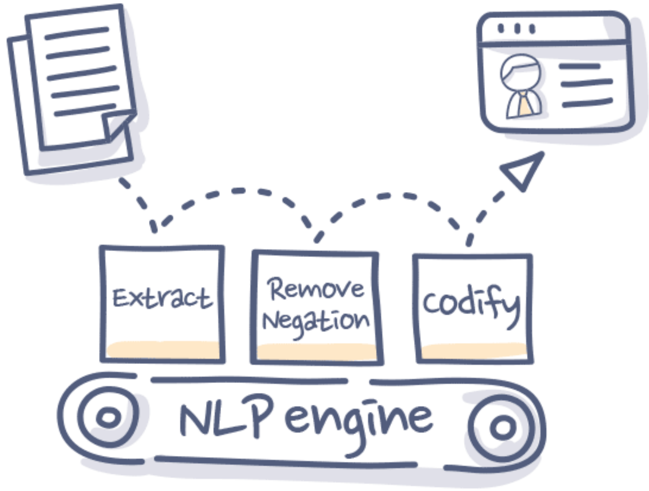
-Friedman C. & Hripcsak G.

PLAYERS IN THIS INDUSTRY

Foresee Medical

- ForeSee Medical's unique combination of machine learning technology and risk adjustment rules delivers industry leading NLP accuracy scores.
- How it works:

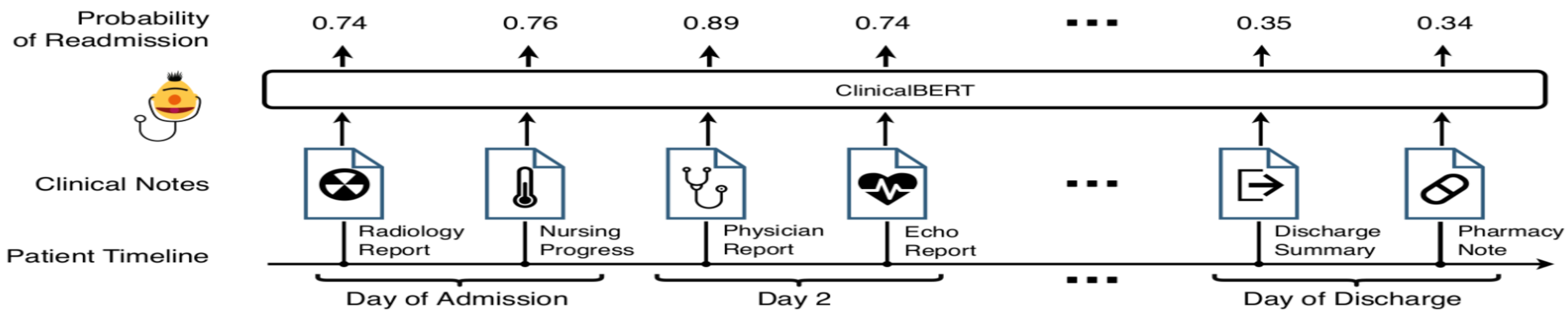


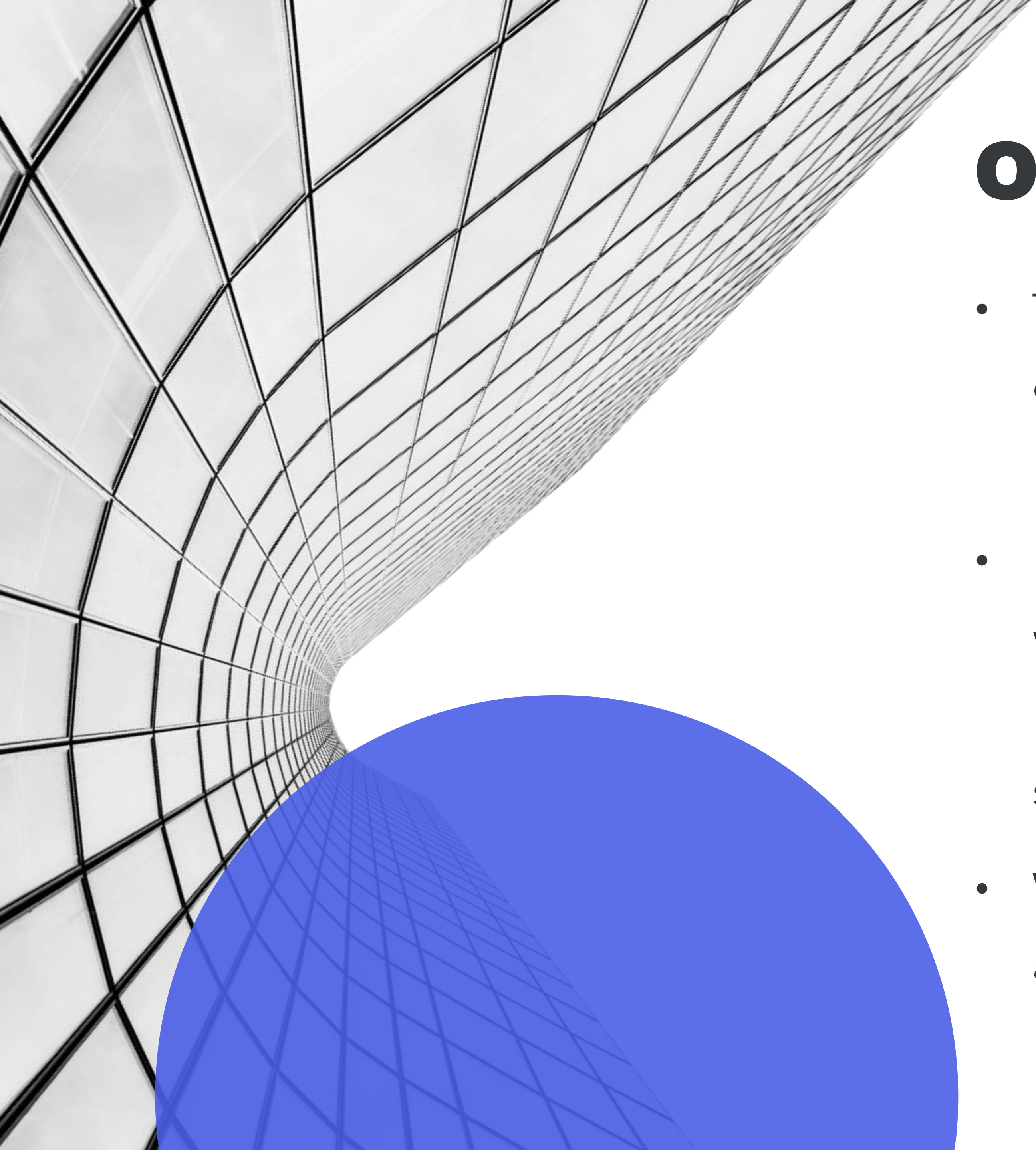


REFERENCE RESEARCH

ClinicalBERT

- ClinicalBERT is a machine learning model that uses clinical notes and electronic health records (EHRs) to predict hospital readmissions.
- The model is based on BERT pre-trained language model and fine-tuned on EHRs to capture specific language used in clinical notes.
- The evaluation of ClinicalBERT on a large dataset of patient records from two hospitals showed that it outperformed several other baseline models in predicting readmissions within 30 days of discharge.
- The research suggests that ClinicalBERT and other machine learning models have the potential to improve healthcare outcomes by helping clinicians identify high-risk patients and prevent hospital readmissions.





OUR PRODUCT

- The healthcare sector being very vast, we decided to focus on solving a specific problem at first.
- Hence, we plan on building a product which would be capable of predicting patient hospital readmission with their discharge summary.
- We believe this would help the patient as well as the hospital to plan its resources.

PRODUCT DEVELOPMENT PLAN

RESEARCH	PLANNING	DESIGN	DEVELOPMENT	LAUNCH
Conduct brief **literature/industry research** to determine similar projects. Learn from those studies' outcomes and differentiate their project than ours.	Plan and document the details of the planned implementation. Get familiar with the **datasets** and carry out transformations and cleansing. Carry out some basic **exploratory data analysis**.	Getting the dataset completely ready after appropriate **cleansing and transformations**. Getting familiar with all the major **patterns and trends** in dataset.	**Construct** the model. Produce concrete **outcomes**. Focus on **performance** of the model.	Deploy the tool on **Streamlit cloud.**

TIMELINE

FEB 6TH — **Team** creation and **project** selection

FEB 27TH — Discuss initial **EDA** findings

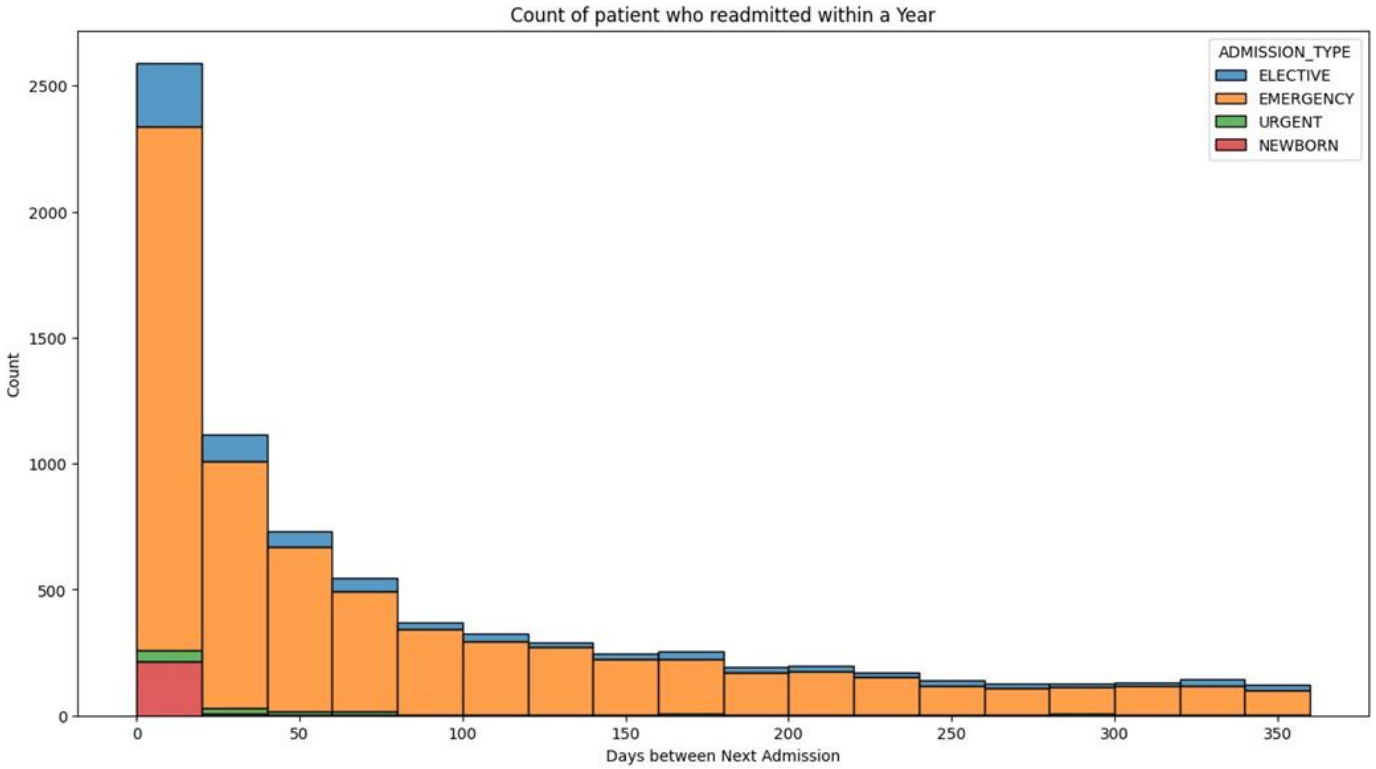
MAR 13TH — Present **Phase-1** presentation, start **design and development** stage

MAR 27TH — Present **Phase-2** presentation, complete **model evaluation**

MAY 1ST — Present **Phase-3** presentation, execution and interpretation, **deploy product**

EDA-1

READMISSIONS WITHIN A YEAR

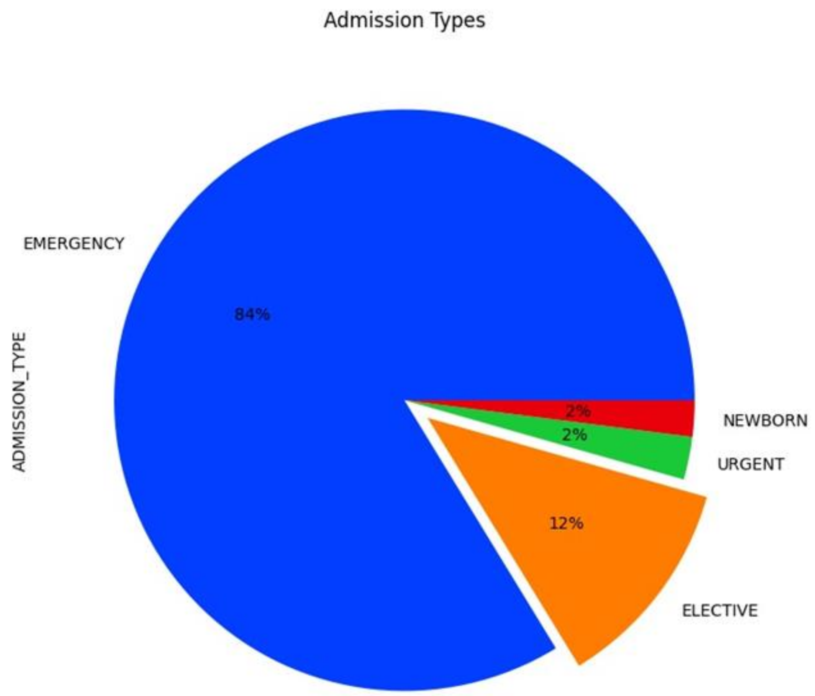


Interpretation:

- The majority of readmissions occur within the first 50 days following discharge.
- Elective admission types have the highest probability of readmission throughout the year, followed by emergency cases.
- Newborn and urgent admission types have the lowest probability of readmission.

EDA-2

UNDERSTANDING ADMISSIONS

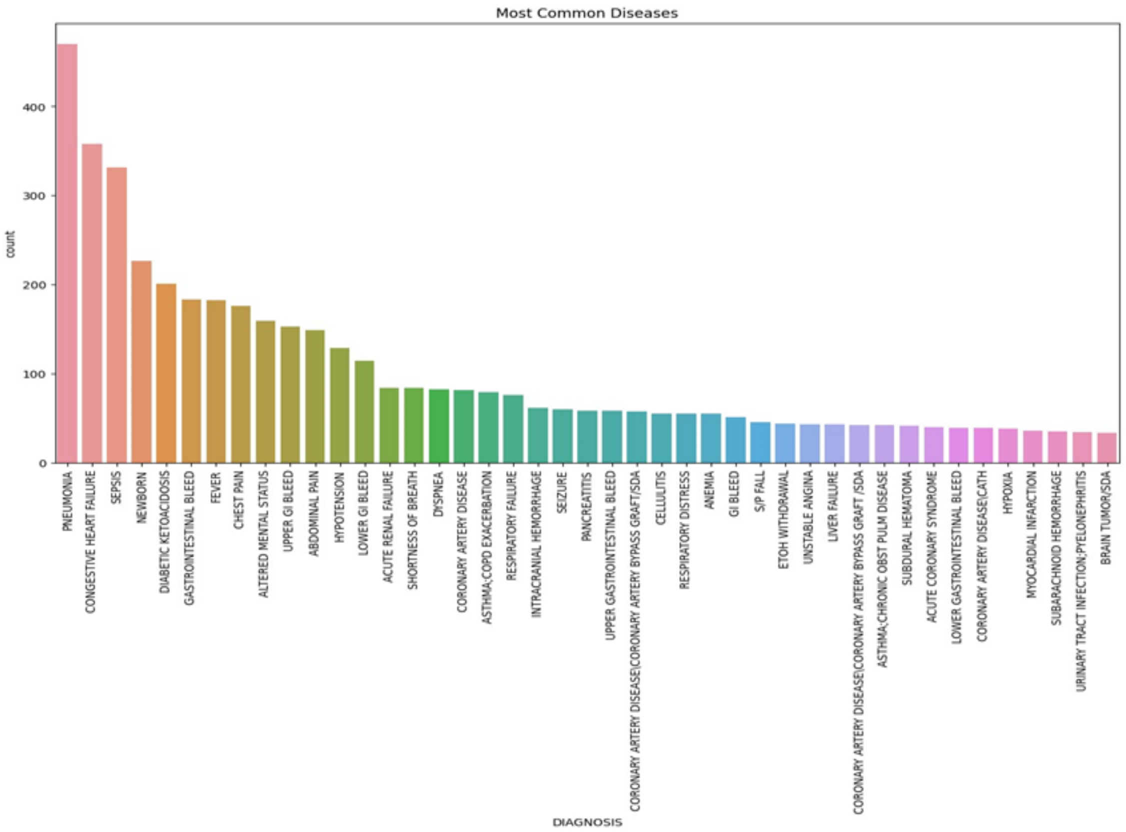


Interpretation:

- Among all admitted patients, 84% were admitted as emergency cases.
- Only 12% of admissions were elective, while urgent and newborn admissions accounted for just 2% each.

EDA-3

UNDERSTANDING DISEASES

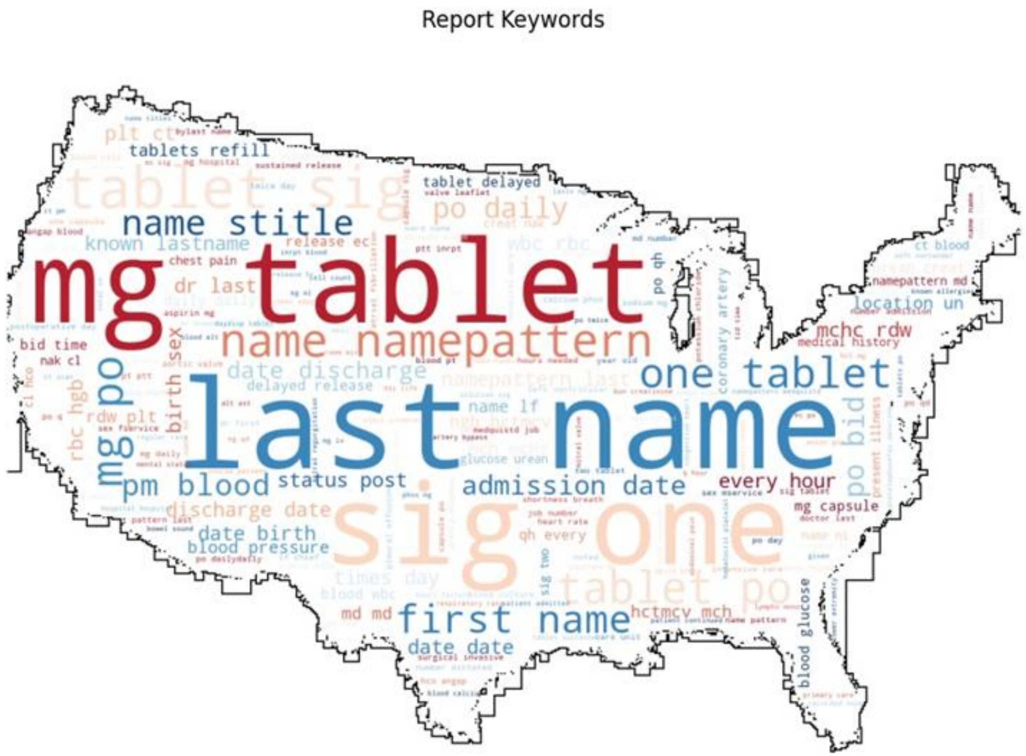


Interpretation:

- Among the top 1% of the most common diseases that result in readmission, Pneumonia, Congestive Heart Failure, and Sepsis are the leading causes, followed by other illnesses.
- This highlights that patients who have been diagnosed with these diseases are more likely to require readmission for further diagnosis and treatment.

EDA-4

REPORT ANALYSIS



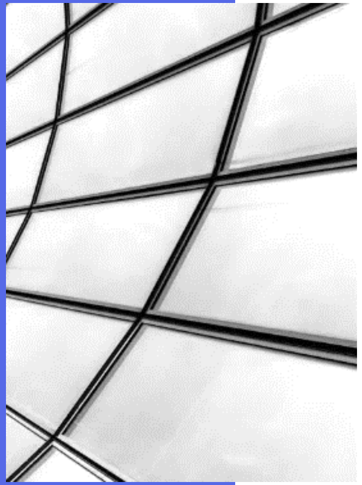
Interpretation:
- Individual patient report keywords identified.
- We get to know insights like the following:
  - What information is present in the reports on a high level?
  - Is that information useful for prediction, for example dosage?
- What features should we focus on?

REFERENCES

- Friedman, C., & Hripcsak, G. (1999). Natural language processing and its future in medicine. Academic medicine : journal of the Association of American Medical Colleges, 74(8), 890–895. https://doi.org/10.1097/00001888-199908000-00012
- Huang, K., Altosaar, J., & Ranganath, R. (2019). Clinicalbert: Modeling clinical notes and predicting hospital readmission. *arXiv preprint arXiv:1904.05342*.

DATA INFO

- MIMIC-III is a relational database consisting of 26 tables (https://physionet.org/content/mimiciii/1.4/).
- Database size is approximately 20GB, where as ADMISSIONS.csv is 12MB and NOTEEVENTS.csv is 3.9GB.
- 2.08 Million records present in NOTEEVENTS.csv



MEET OUR TEAM

Suhetu Ring

Student, UMBC

Harsha Vanga

Student, UMBC

Harshit Shrimali

Student, UMBC

THANK YOU

Team MedBots