# Medical Text Analysis

Prediction of Diabetes and it’s many Co-morbidities

# Table of contents


01 **Objectives**
Why do we need to analyse medical text

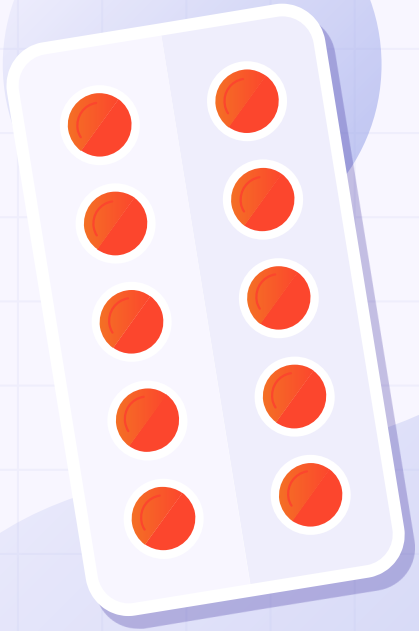
02 **Methodology**
What data and methods we used to predict

03 **Results analysis**
The results of our analysis

04 **Conclusions**
The predictions and future direction

# Goal

Clean and process the raw unstructured text data to predict (Yes or No) if a person has diabetes and it's many comorbid disorders



# Why?

**1**

Analyzing medical text through NLP techniques in this project facilitates disease prediction and co-morbidity assessment.

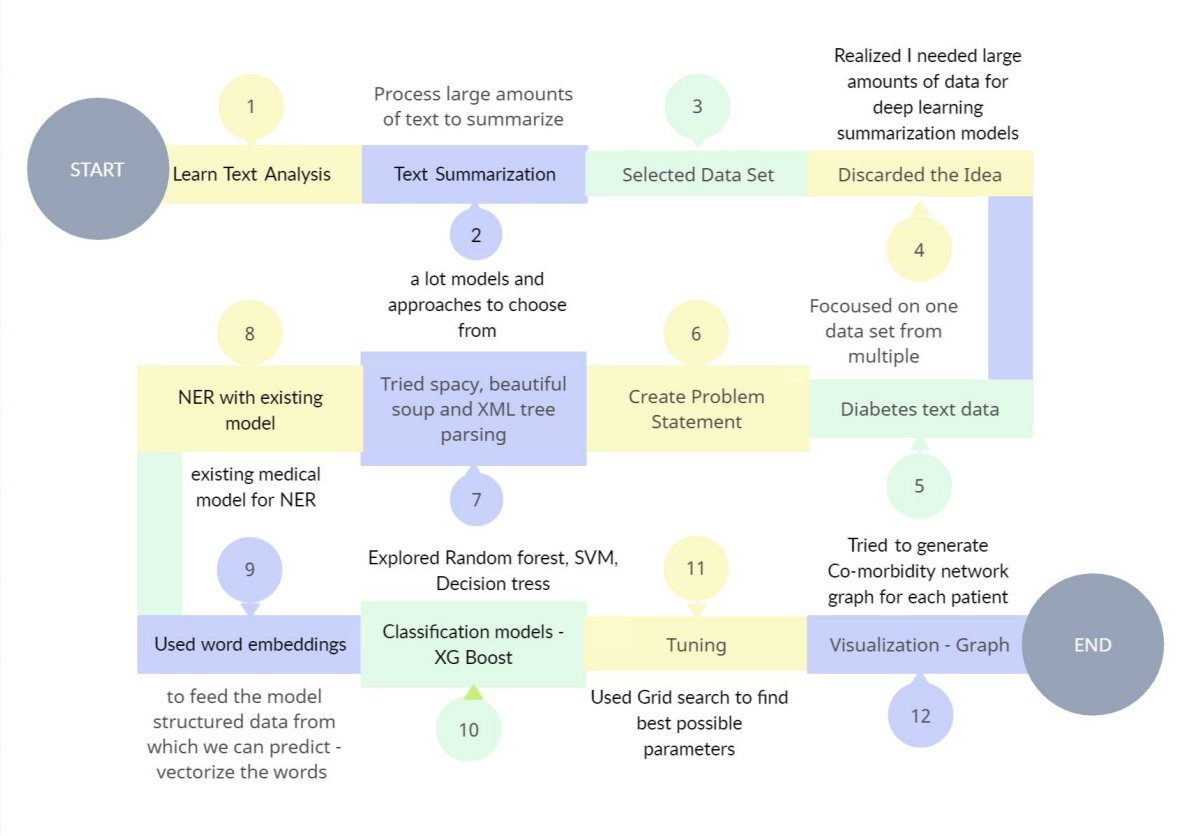
**2**

One person or medical professional might miss all the possible comorbidities associated with diabetes

**3**

While structured values and categories offer easy analysis and results, most information is recorded in the medical text

# Knowledge Process



# Methods

1. Parse the XML file and make a data frame
2. Remove stopword, Tokenize text
3. Apply Named Entity Recognition with existing hugging face model
4. Extract Word Embeddings using Clinical BERT model
5. Use XGBoost Classifier to predict disease outcome
6. Use GridSearch to get the parameters for best possible accuracy

# Accuracy

60%

**Congestive Heart Failure**

74%

**Diabetes**

The model with it's best parameters predicts the outcome of Diabetes with 74% accuracy

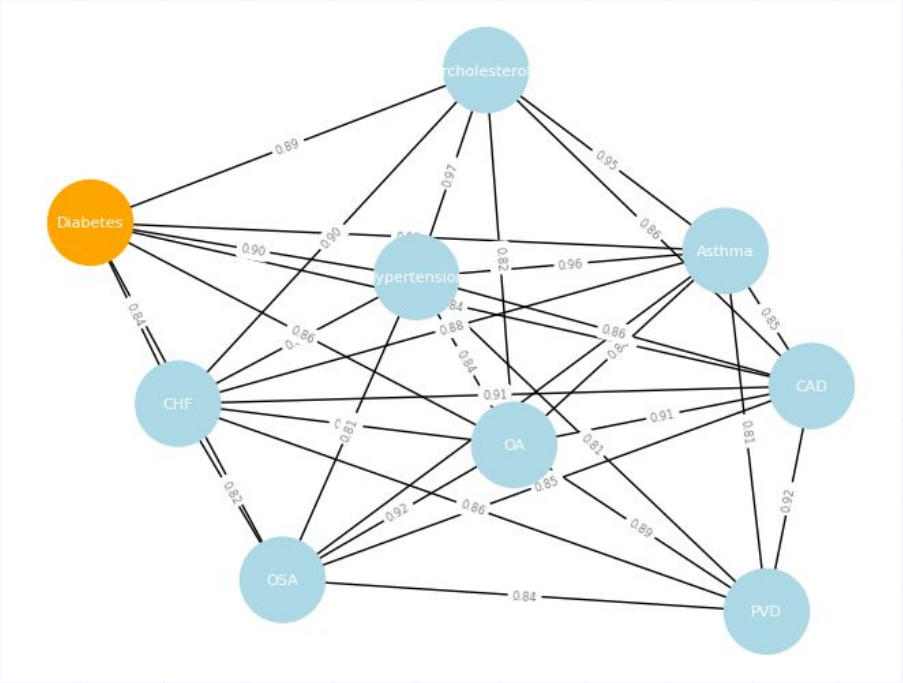
75%

**Osteoarthritis**

# Direction and final goal

Can be used to map all disorders

The values of how closely the words are related




A Chart can be generated for each patient

When medical text is provided - output their comorbidity network graph

# Conclusions

By using the medical text we can with reasonable accuracy predict if a person has diabetes or not. Accuracy can be improved with more data and better medically trained models

# Thank You!

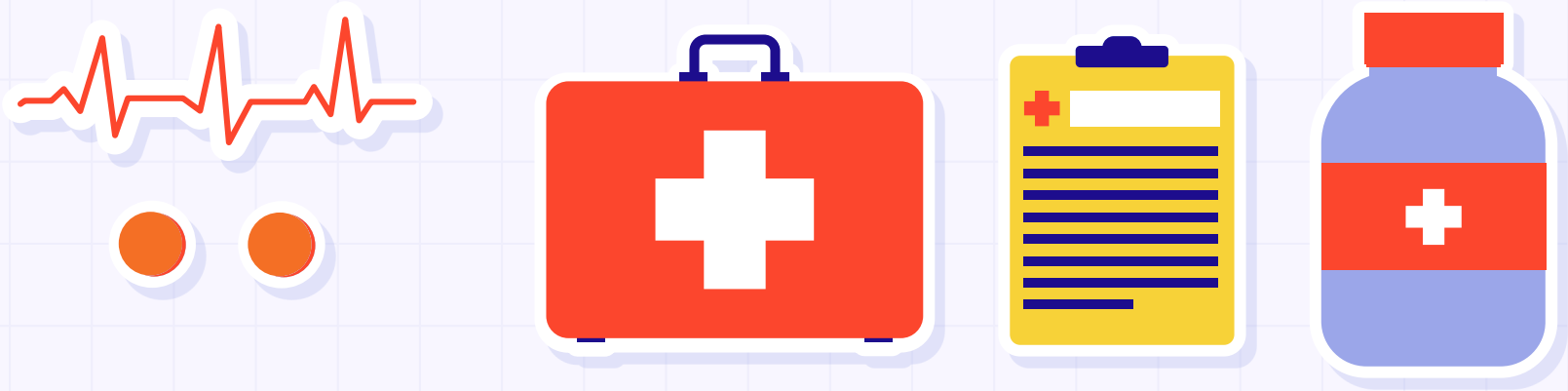
## Do you have any questions?

sai.p.lakkireddy.gr@dartmouth.edu

## Github link (Request access)

https://github.com/saipriya0209/Project_Diabetes_Clinical_Text_Analysis_QBS101_5.git

CREDITS: This presentation template was created by **Slidesgo**, including icons by **Flaticon** and infographics & images by **Freepik**

# Resources

- Harvard n2c2 NLP Data Sets
- Stack Overflow
- Chat GPT
- Hugging Face
- Özlem Uzuner, PhD, Recognizing Obesity and Comorbidities in Sparse Data, Journal of the American Medical Informatics Association, Volume 16, Issue 4, July 2009, Pages 561–570, https://doi.org/10.1197/jamia.M3115