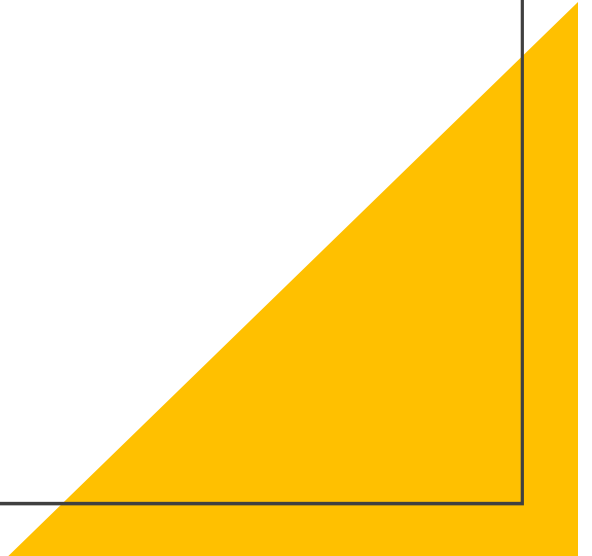


Stroke Prediction

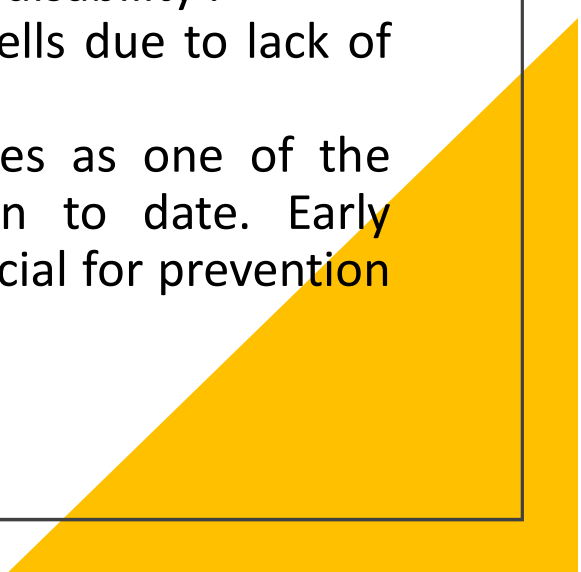
CMPE-255 : Short Story Assignment

By

Mamatha Guntu

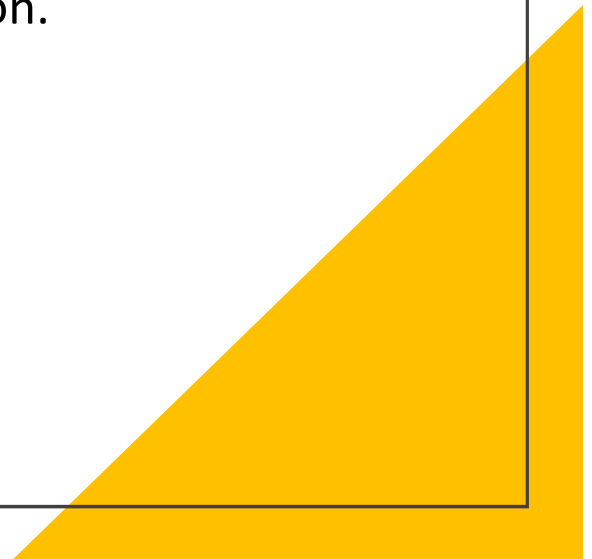


Introduction : What is Stroke ?

- Cerebrovascular accidents, also known as strokes, are the second leading cause of death worldwide and the third leading cause of disability .
 - Stroke is defined as the sudden death of some brain cells due to lack of oxygen and, in many cases, it is asymptomatic
 - The World Health Organization (WHO) declares strokes as one of the growing crises that is receiving very little attention to date. Early awareness of having this illness has proved to be beneficial for prevention and treatment.
- 
- A large yellow triangle is positioned in the bottom right corner of the slide, pointing towards the top right.

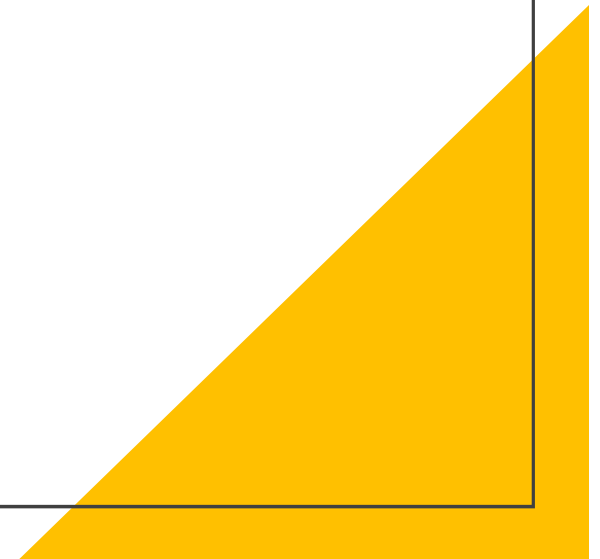
What is the problem we are trying to solve ?

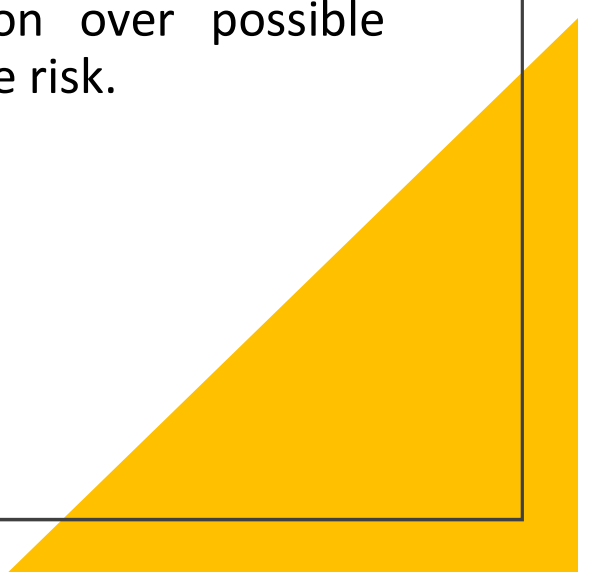
- Determine whether or not a patient will have a stroke in the next year automatically using her/his historical medical information.

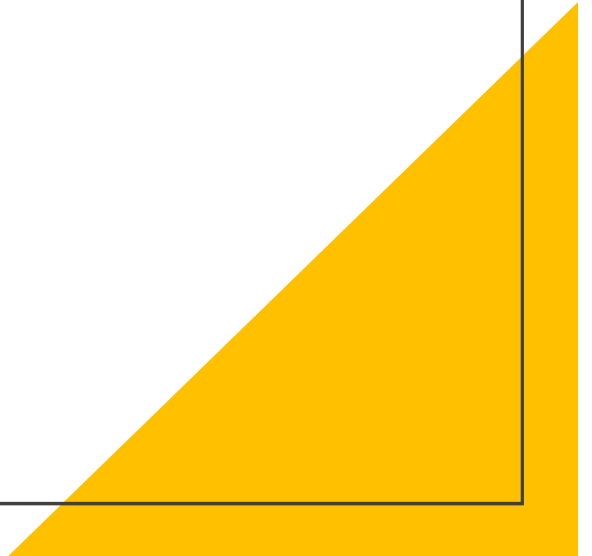


What are the methods that can solve the problem ?

- CHADS2 is a score metric to evaluate stroke risk prediction proposed by Gage.
 - C – congestive heart failure
 - H – Hypertension
 - A – aged 75 or older
 - D - Diabetes
 - S2 – who had stroke in the past

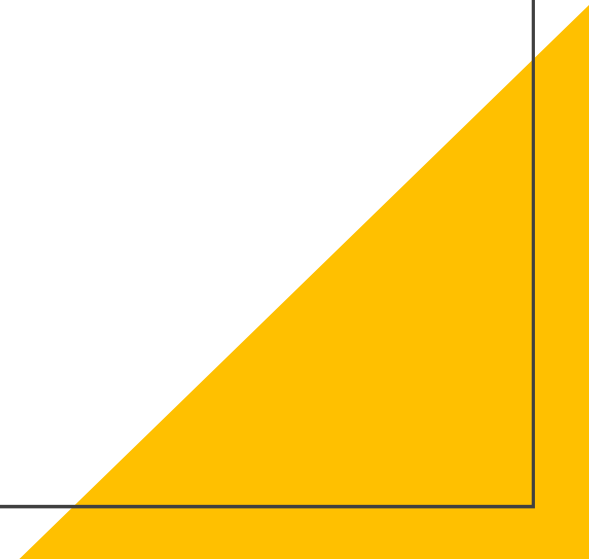


- Bayesian Rule Lists (BRL) to develop an interpretable model to predict stroke risk within a year for patients diagnosed with atrial fibrillation. The method uses decision lists, which consist of a series of “if . . . then . . .” statements, then it uses a generative model called Bayesian Rule Lists which automatically produces a posterior distribution over possible decision lists allowing inferences to be made about stroke risk.
- 
- A large yellow right-angled triangle is positioned in the bottom right corner of the slide, extending from the bottom edge and the right edge of the main content area.

- Teoh also proposed a method to predict stroke risk. In his work, a Recurrent Neural Network (RNN) was used in combination with a custom loss function. The model uses all available data for a patient, such as exam results and diagnosis structured like a time series for the RNN. Then, fully-connected layers are applied to predict the class
 - The final method by Weng is to work on routine clinical data and machine learning techniques.
- 

What kind of data is required for the analysis ?

- Patient checkup data
- Patient demographics
- Patients disease history



How to shape the data ?

- Handle Missing values

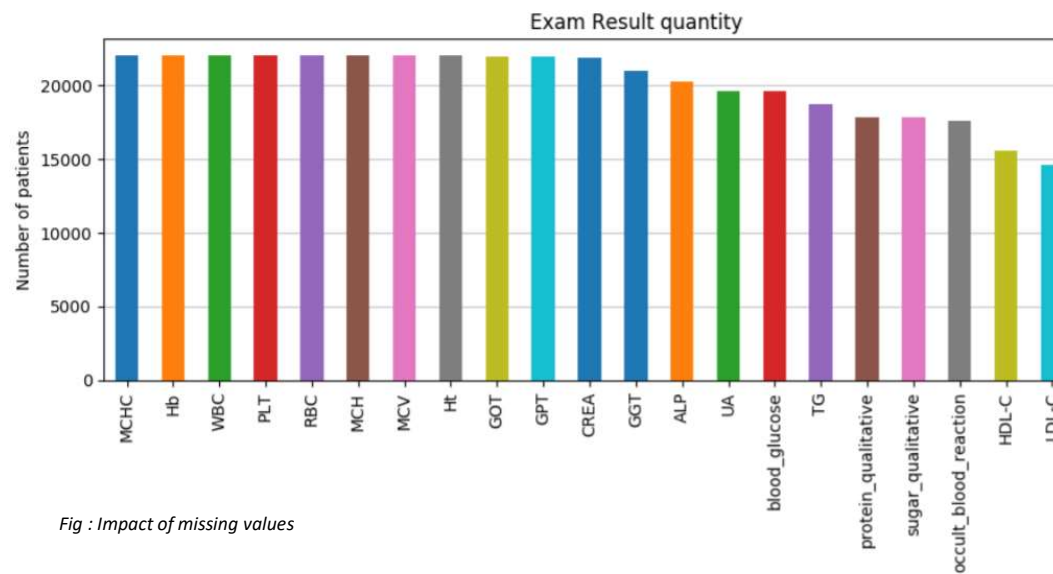
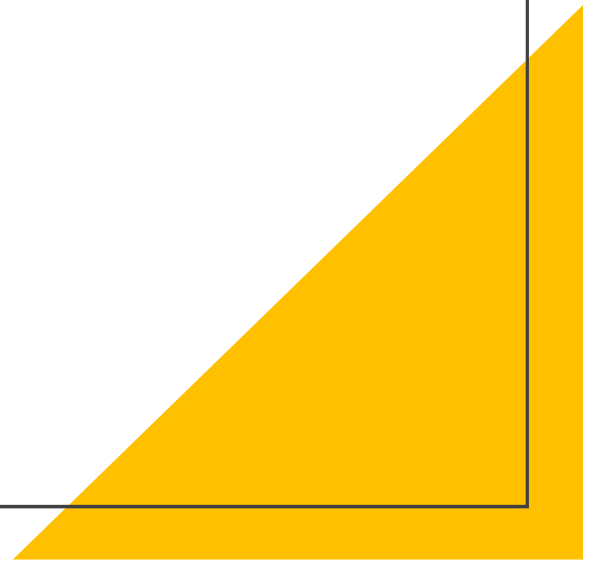


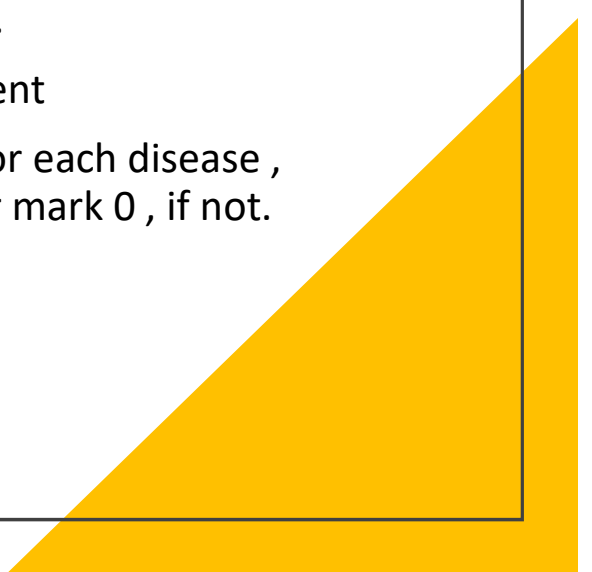
Fig : Impact of missing values

How to model and fit the data ?

- Embedding
- Classifier and rules



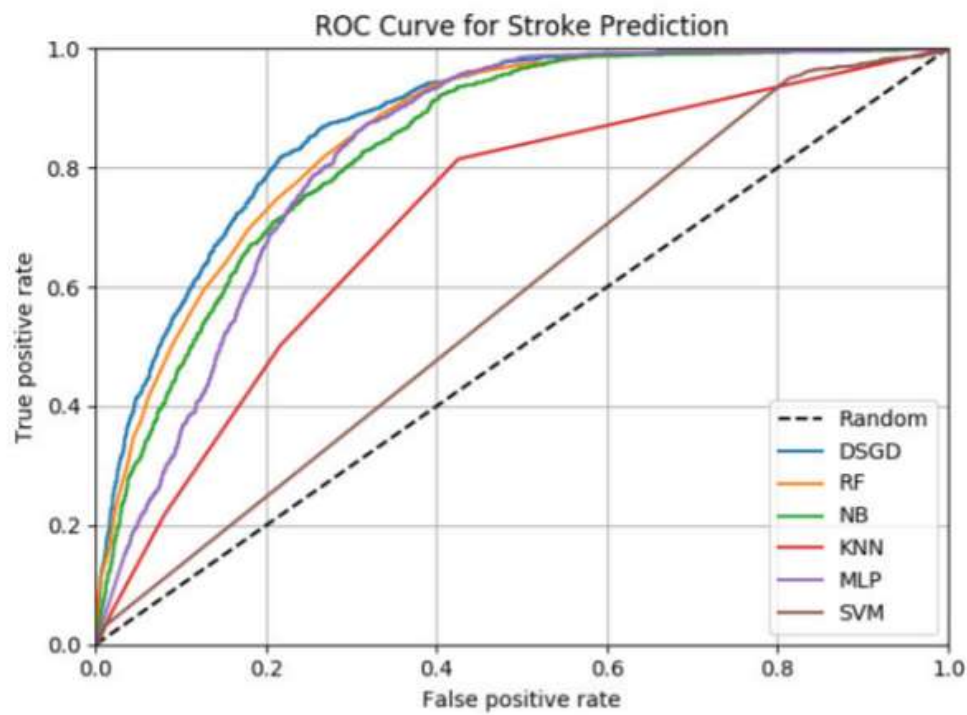
Embedding

- Process of representing an observation or record as a structured data, typically, a vector.
 - **Drawback:** Multiple exam result records exist for the same patient .
 - **Solution:** Use most recent and current condition record of the patient
 - **Approach :** consider the disease history , have a separate column for each disease , mark 1- if the patient was diagnosed with the disease in the past or mark 0 , if not.
- 
- A large yellow triangle is positioned in the bottom right corner of the slide, pointing towards the top right.

Classifier and Rules

- This model is rule-based and allows to generate rules automatically or to define custom rules according to data; the model can handle missing information which is a requirement according to the chosen embedding; and the model is interpretable which means that it can give an explanation of any prediction.
- Rules
 - For exams that are qualitative and semi quantitative, and for past history of diseases , create a rule for each possible outcome
 - Define the cut off values between normal and abnormal values for each attribute. These cut of values are called reference interval. Activate the rule based on the cut off values.
 - Create rules that activate when pair of attributes are outside the range , this will help search for more complex relationships in data.

Data visualization



ROC curve for various methods in the stroke risk problem

Model	Accuracy	AUC ROC
DSGD	0.854	0.875
RF	0.849	0.861
NB	0.618	0.838
KNN	0.794	0.717
MLP	0.821	0.813
SVM	0.820	0.574
Majority	0.793	0.500

Validation

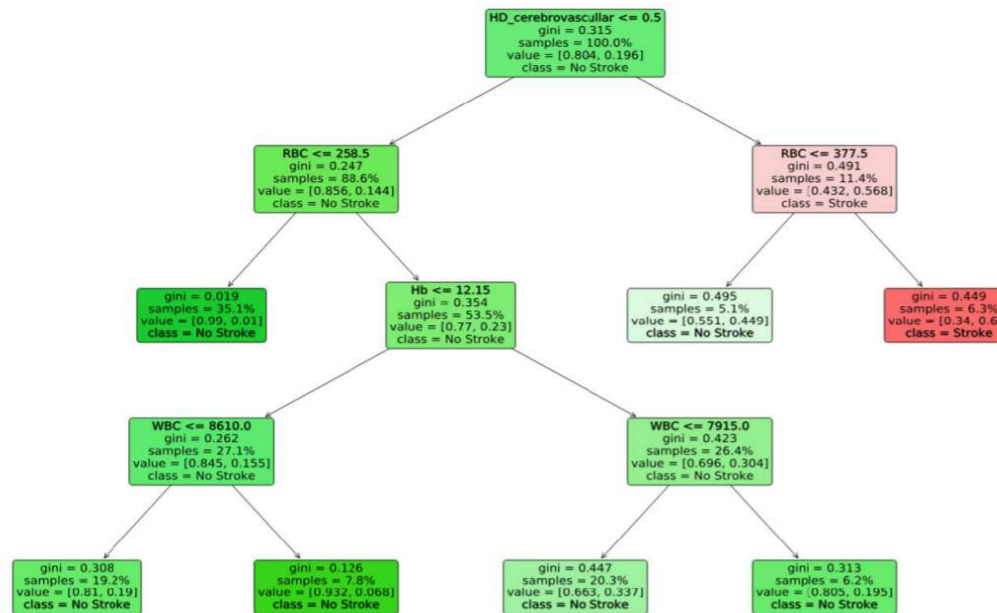


Figure: Decision tree for stroke risk validation

References

- <https://ieeexplore-ieee-org.libaccess.sjlibrary.org/stamp/stamp.jsp?tp=&arnumber=9306826>

Thank You !