

# HEART MURMUR- PREDICTIVE MODEL

Presented by: Kartik Sojitra

# INTRODUCTION

- **INTRODUCTION**

- “HEART MURMUR” IS ONE OF THE WORLD’S LARGEST DISEASE AS OF NOW.
- IT IS LEFT UNDIAGNOSED IN 80% OF CASES.
- A DETAILED ANALYSIS IS NEEDED TO IDENTIFY MODEL PREDICTS HEART MURMUR ACROSS AGE GROUPS
- THE PROJECT ANALYSES THE HEART VALVE MEASUREMENTS USING ECHO CARDIOGRAPH REPORTS OF 5009 PATIENTS AT PRINCESS MARGRET HOSPITAL, IN PARTNERSHIP WITH THE UHN ECHOCARDIOGRAPHY LAB.
- THE REPORT ANALYSES THE DATASET USING DETAILED PROCESS OF MACHINE LEARNING, WHICH HELPS TO IDENTIFY AND ADDRESS THE PROBLEM.

# PROBLEM STATEMENT

“Will the report be able to identify and develop a predictive model which can determine Children with Congenital heart defects and Adults with Heart valve defects using machine learning algorithm?”

The answers for these questions will help the audience to take it further developing a business model which can help to creating a portable device which can predict and identify Heart murmur.

Heart valve defects can be seen at the time of birth with new born babies and also it is seen in adults as well. With this project, the algorithm will find out a prediction model to diagnose heart defects in other patients across the age group



# KEY FEATURES

**Will the model be able to identify and predict Heart Murmur in patients at different age group?**

Yes, A Classification report with the confusion matrix, learning curve, box-plots are the metrics which will be used to evaluate the algorithm to find out accurate responding model. The model will look for output variable classified as 0-Normal patients, 1- Congenital heart defect-Babies, 2- Heart Valve Defect-Adults.

---

Congenital heart defects have been left undiagnosed which have results existence of heart murmur in 40-45% of Children.

---

Heart valve defects in adults are also undiagnosed which results heart murmur in 10% of Adults.

---

Early diagnosis of heart defects helps to reduce 90% of heart episodes. So, by analysing the data, the project can identify a pattern using classification model, to predict heart valve defects

**Is it that important to identify Heart Murmur?**

**Accuracy above 90%, will that help developing a perfect predictive model?**

Yes, the metrics which the report looks at are, Precision, Recall and Accuracy. As it is medical data, higher accuracy always gets you're the better model with less errors.

# CONSTRAINTS

- AGE GROUP IS NOT MENTIONED WHICH MAY BRING DOWN ACCURACY
- LACK OF SOURCES TO ADD MORE INFORMATION
- AS WE DEVELOPING A PREDICTIVE MODEL IDENTIFYING THE HEART DEFECT, THE ACCURACY MUST BE MORE THAN 90% AT LEAST. IF THE PATIENTS INVOLVED HAVE BEEN USED DIFFERENT DIAGNOSING TECHNIQUES MEASURING THE READINGS, IT MAY CAUSE IMBALANCED VARIANCE- BIAS. THIS FURTHER WILL RESTRICT TO HAVE BETTER ACCURACY.

# DATASET-CLEANING

- THE DATASET HAVE BEEN EXPLORED TO CHECK THE DISTRIBUTION FOLLOWED BY NOTICED AND REMOVED NULL VALUES ALONG WITH OUTLIERS.
- THE SAME WAS THEN CROSS VALIDATED .
- THE SIZE OF THE DATASET WAS IMBALANCED AND BALANCED USING SMOTE TECHNIQUE WHICH SYNTHETICALLY CREATES ARTIFICIAL DATAPOINTS.
- THE DATASET WAS EXPLORED TO IDENTIFY IF THERE ANY INTERCORELATION.

# EDA



Data explorations



Data balancing



Data distribution



Data feature selection



Data intercorrelation



Preparing data for running with model

# MODEL ANALYSIS



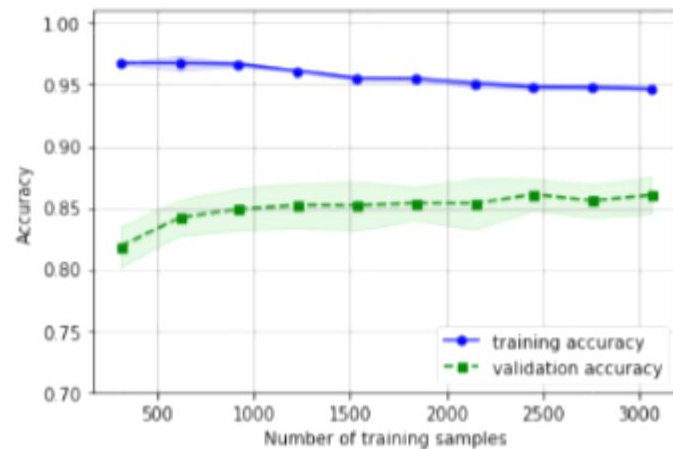
The model will analyzed with learning curve, confusion matrix and classification report.



Precision , recall and accuracy will be considered as important metrics to compare models.



SVM Learning Curve



# FIRST MODEL PREFERRED

- SUPPORT VECTOR MACHINE
- SUPERVISED MODEL OPERATES FORMING A HYPERPLANE TO SPLIT BINOMIAL OUTPUTS USING SUPPORT VECTOR. DISTANCE BETWEEN SUPPORT VECTOR NEARER TO THE HYPERPLANE IS CALLED MARGIN. HIGHER THE MARGIN BETTER THE MODEL.

```
[[237 21 30]
 [ 22 244 17]
 [ 13 15 253]]
```

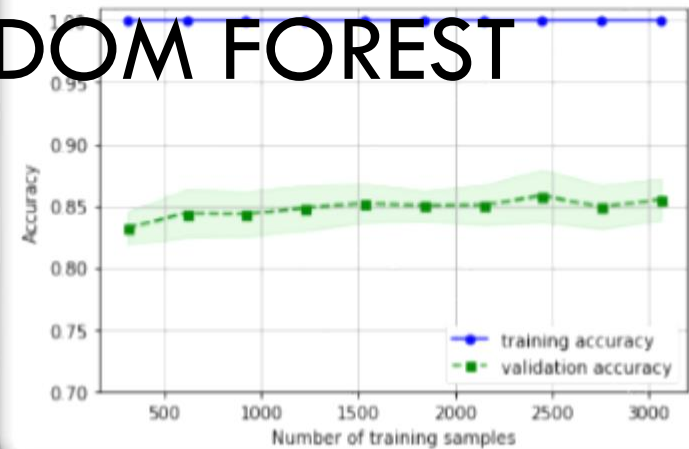
	precision	recall	f1-score	support
0	0.87	0.82	0.85	288
1	0.87	0.86	0.87	283
2	0.84	0.90	0.87	281
accuracy			0.86	852
macro avg	0.86	0.86	0.86	852
weighted avg	0.86	0.86	0.86	852

```
NestedCV Accuracy(weighted) :0.86 +/-0.01
NestedCV Precision(weighted) :0.86 +/-0.01
NestedCV Recall(weighted) :0.86 +/-0.01
```

# ALTERNATIVE MODEL-RANDOM FOREST

- EACH OF THE DECISION TREE GIVES A BIASED CLASSIFIER (AS IT ONLY CONSIDERS A SUBSET OF THE DATA). THEY EACH CAPTURE DIFFERENT TRENDS IN THE DATA. THIS ENSEMBLE OF TREES IS LIKE A TEAM OF EXPERTS EACH WITH A LITTLE KNOWLEDGE OVER THE OVERALL SUBJECT BUT THOROUGH IN THEIR AREA OF EXPERTISE.

RandomForest Learning Curve



```
[[221 32 35]
 [ 18 253 12]
 [ 7 14 260]]
```

	precision	recall	f1-score	support
0	0.90	0.77	0.83	288
1	0.85	0.89	0.87	283
2	0.85	0.93	0.88	281
accuracy			0.86	852
macro avg	0.86	0.86	0.86	852
weighted avg	0.86	0.86	0.86	852

```
NestedCV Accuracy(weighted) :0.85 +/-0.01
NestedCV Precision(weighted) :0.86 +/-0.01
```

# ENSEMBLE VOTING MODEL

The model combines prediction of all models and predicts the model which gets more vote as below,

Voting Model

RandomForestClassifier 0.85

SVM-0.86

XGBClassifier 0.86

VotingClassifier 0.86

# CONCLUSION

Questions	Need of the question
<b>Will the model be able to identify and predict Heart Murmur in patients at different age group?</b>	Yes, Random forest being an ensemble model, can predict with a accuracy of 86% .However, the issue of not getting target 90% will be due to lack of features
	Yes, it is important and the model can identify upto 86% accuracy
<b>Is it that important to identify Heart Murmur?</b>	
<b>Accuracy above 90%, will that help developing a perfect predictive model?</b>	Random forest can get a accuracy above 90% if constraints and relevant features are taken care.

# RECOMMENDATIONS

- THE PROJECT TEAM RECOMMENDS RANDOM FOREST AS A PREDICTIVE MODEL.
- BEFORE TAKING INTO IT, THE TEAM RECOMMENDS TO ADD SOME MORE ADDITIONAL FEATURES WHICH CAN CONTRIBUTE MORE TO THE OUTPUT AND GET ACCURACY MORE THAN 90%.
- PATIENTS GENDER, RACE AND AGE GROUP DETAILS NEEDED.

