

FETAL HEALTH CLASS MODELING

Presented by: Bayan Ali – Eman Albarakah – Yaqeen alhawaj



CONTANTS

Introduction

Algorithm

Workflow

Challenges

Data overview

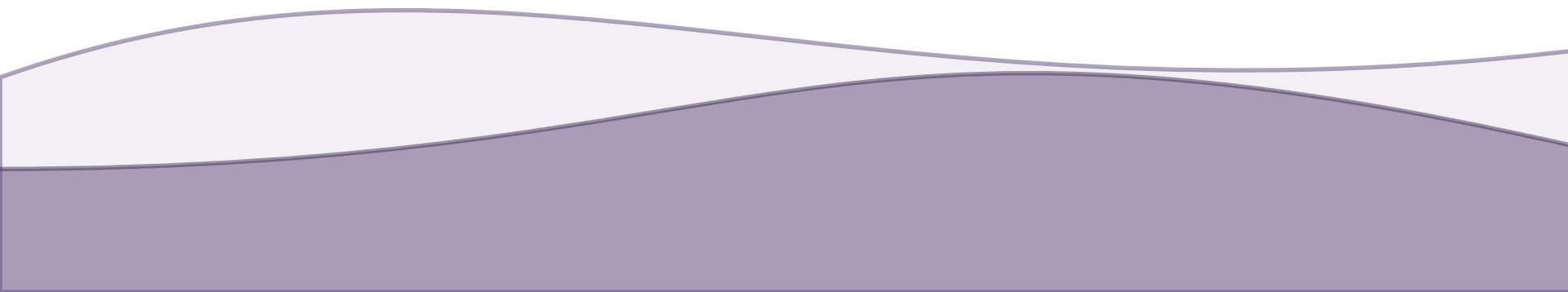
Tools

EDA

Conclusion

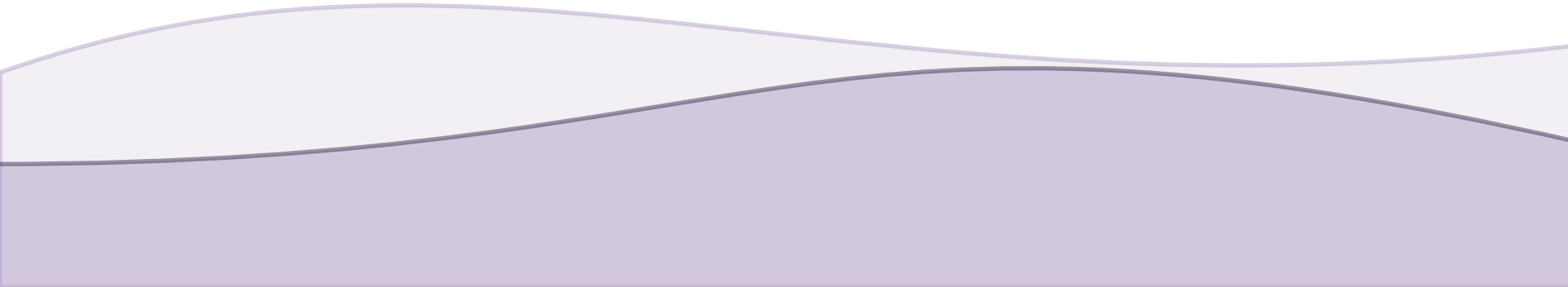
○ Introduction

- Prevent child mortality
- Not everyone has access to technology
- CTGs scans are cost effective and widely spread
- Eliminate erroneous surgical intervention

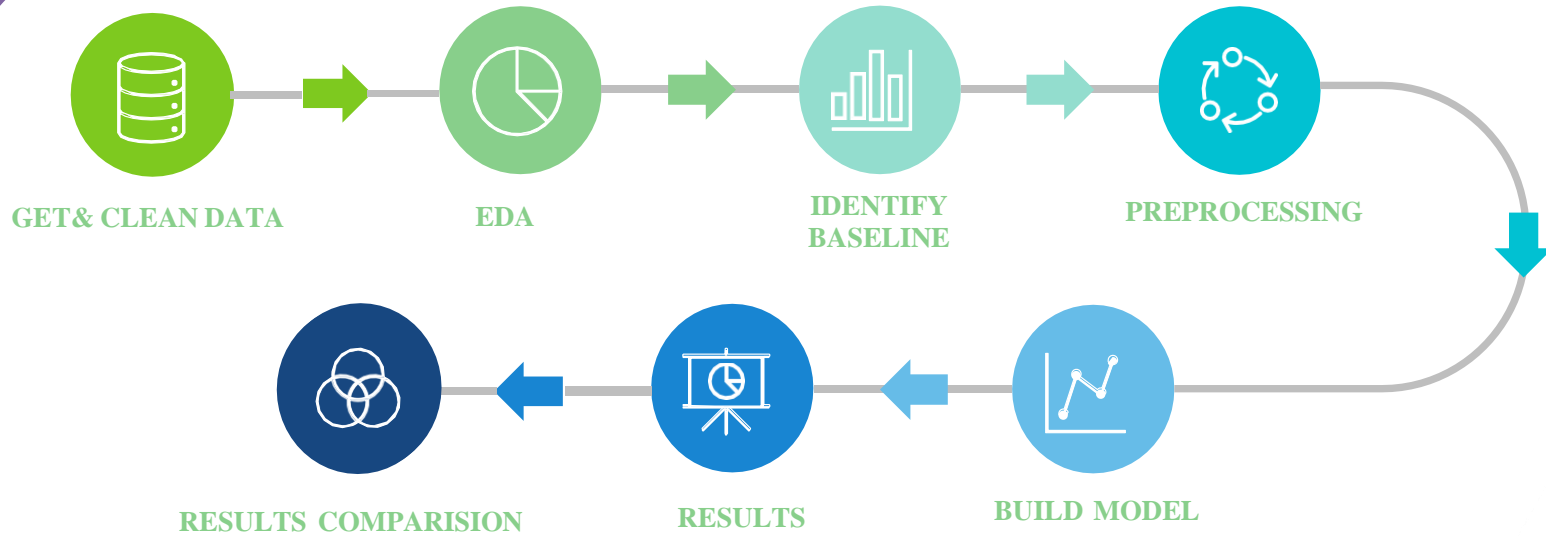


○ Business Problem

- CTG scans are currently interpreted via visual analysis by the physician and reading errors may increase fetal health risk.



Workflow



Data Overview

Kaggle

It has 2126 rows and 22 columns

Feature	Description
baseline value	Baseline Fetal Heart Rate
accelerations	Number of accelerations per second
fetal movement	Number of fetal movements per second
uterine contractions	Number of uterine contractions per second
light decelerations	Number of LDs per second
severe decelerations	Number of SDs per second
prolongued decelerations	Number of PDs per second
abnormal short-term variability	Percentage of time with abnormal short-term variability

- Exploring Data
- Solve Outliers
- Duplicated



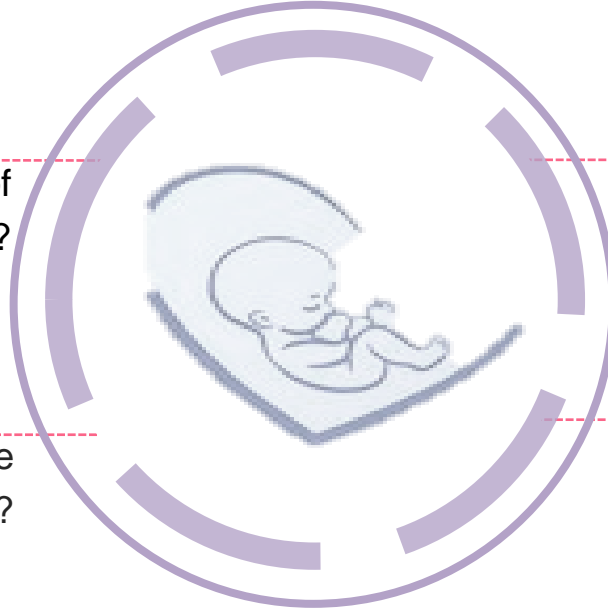
Business Questions

1-what is majority class of fetal health?

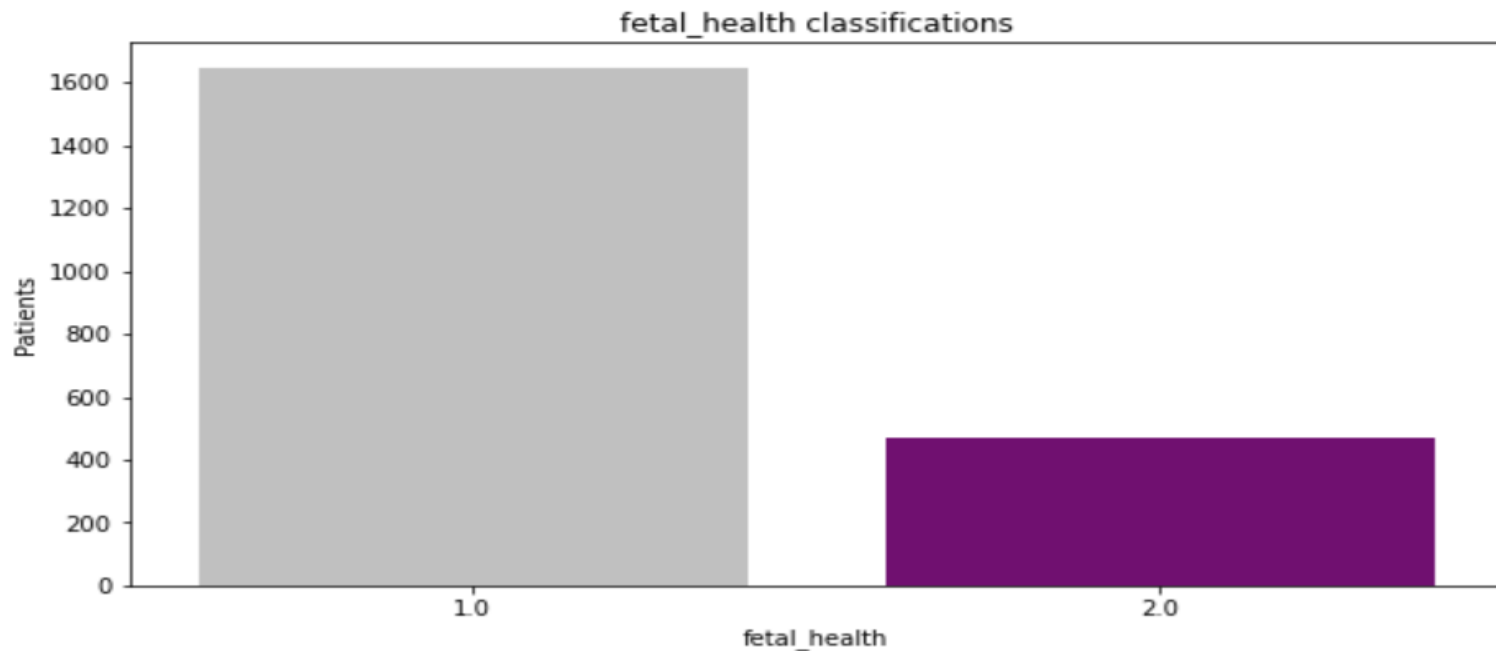
3-What features affect the model?

2-What is the relationship between Baseline Fetal Heart Rate and uterine contractions for Pathological fetal?

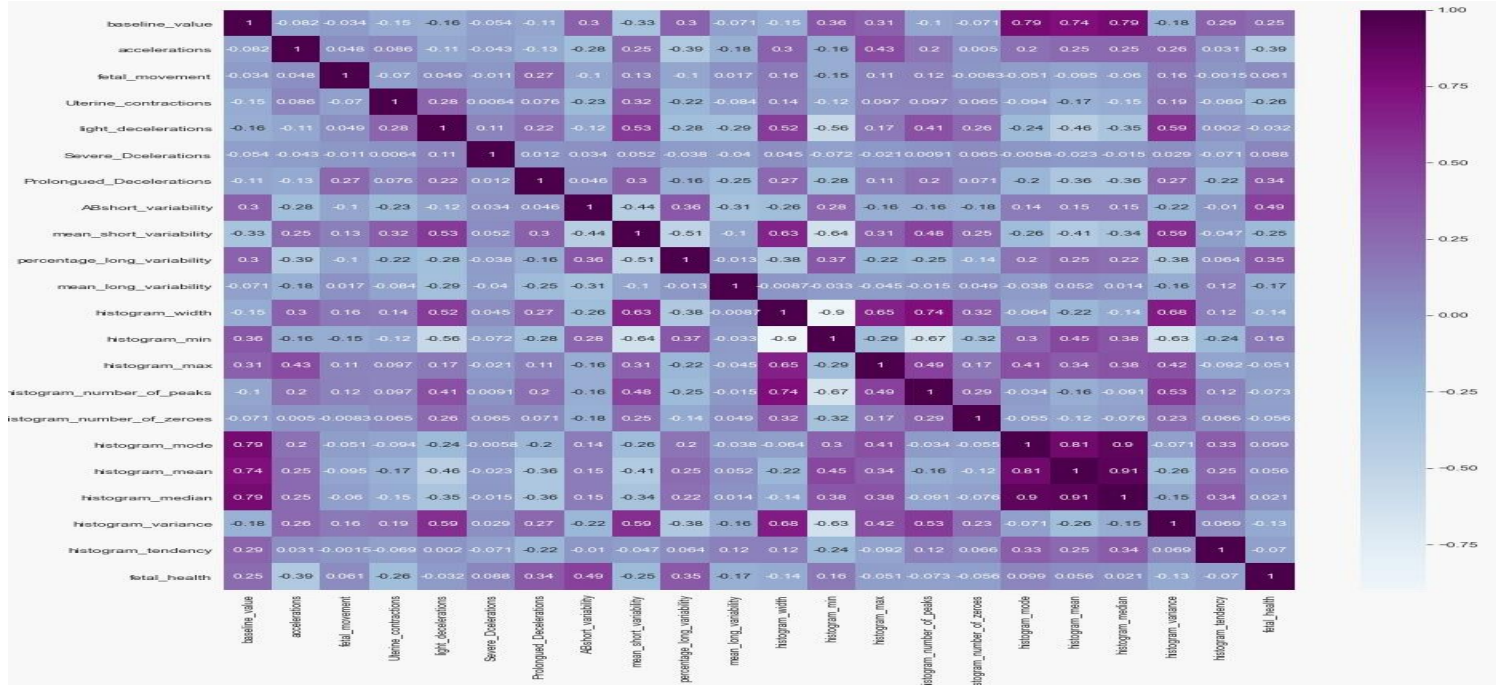
4-Which model makes the best predictions of fetal health class?



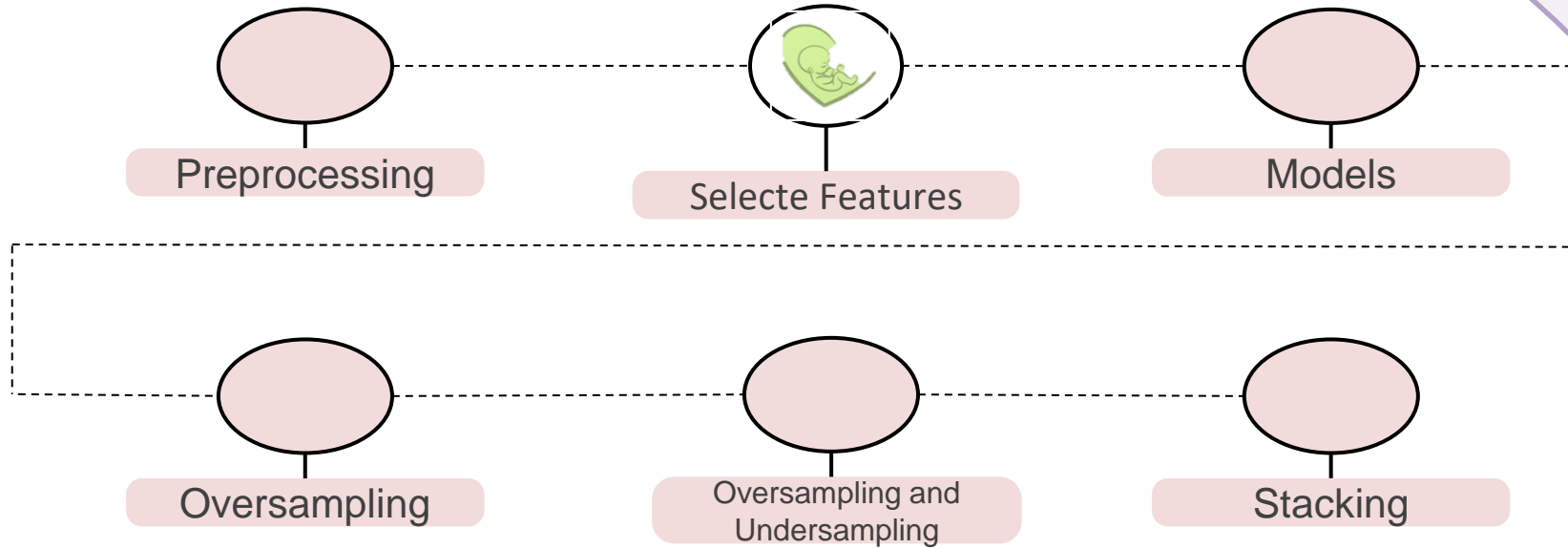
EDA Visualization



EDA Visualization

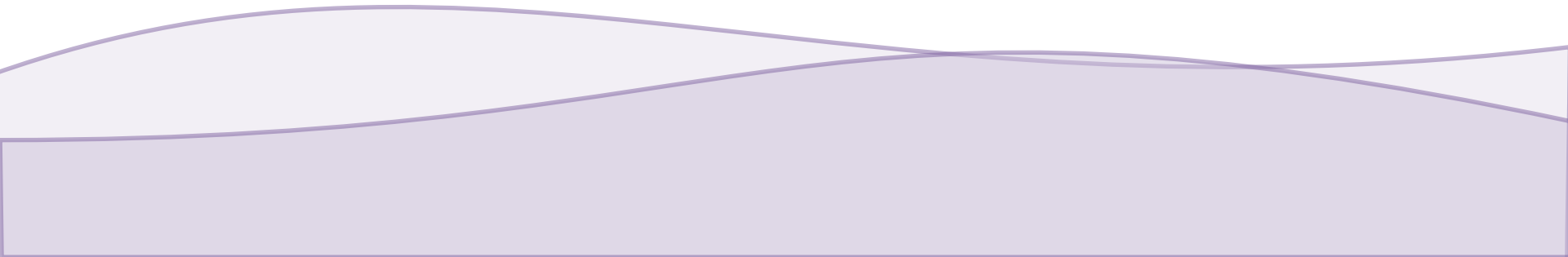


Algorithm



Selected Features

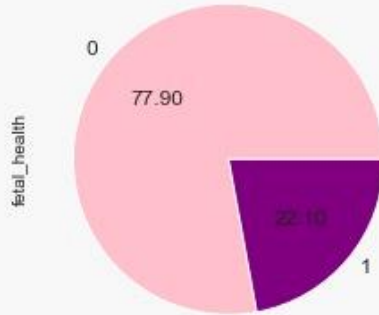
- ☐ Baseline value
- ☐ Accelerations
- ☐ AB short variability
- ☐ Mean short variability
- ☐ Percentage long variability
- ☐ histogram width
- ☐ Histogram min
- ☐ Histogram max



SMOTE

We found that we have a problem of imbalanced data set, namely that target of 1 is less than number of 0.

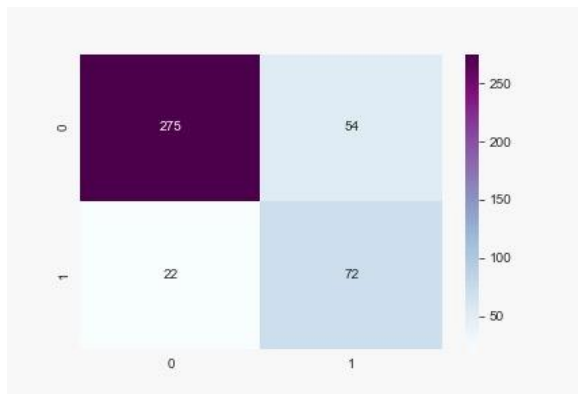
Imbalance y



Balance y



Naive Bayes



Accuracy of Naive Bayes model: 82.03309692671394

	precision	recall	f1-score	support
0	0.93	0.84	0.88	329
1	0.57	0.77	0.65	94
accuracy			0.82	423
macro avg	0.75	0.80	0.77	423
weighted avg	0.85	0.82	0.83	423

Logistic Regression



Accuracy of Logistic Regression: 85.1063829787234

	precision	recall	f1-score	support
0	0.90	0.91	0.90	329
1	0.67	0.65	0.66	94
accuracy			0.85	423
macro avg	0.79	0.78	0.78	423
weighted avg	0.85	0.85	0.85	423

Random Forest



Accuracy of Random Forest: 92.43498817966903

	precision	recall	f1-score	support
0	0.92	0.99	0.95	329
1	0.94	0.70	0.80	94
accuracy			0.92	423
macro avg	0.93	0.84	0.88	423
weighted avg	0.93	0.92	0.92	423

Extreme Gradient Boost



Accuracy of Extreme Gradient Boost: 88.88888888888889

	precision	recall	f1-score	support
0	0.88	0.99	0.93	329
1	0.93	0.54	0.68	94
accuracy			0.89	423
macro avg	0.91	0.77	0.81	423
weighted avg	0.89	0.89	0.88	423

Decision Tree

Accuracy of DecisionTreeClassifier: 89.12529550827422

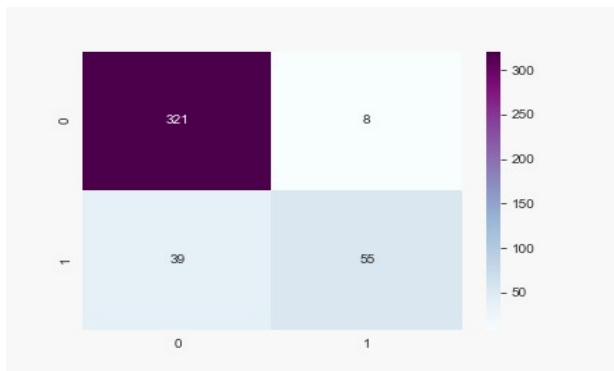
	precision	recall	f1-score	support
0	0.95	0.91	0.93	329
1	0.73	0.82	0.77	94
accuracy			0.89	423
macro avg	0.84	0.87	0.85	423
weighted avg	0.90	0.89	0.89	423



K-Nearest Neighbor

Accuracy of K-NeighborsClassifier: 88.88888888888889

	precision	recall	f1-score	support
0	0.89	0.98	0.93	329
1	0.87	0.59	0.70	94
accuracy			0.89	423
macro avg	0.88	0.78	0.82	423
weighted avg	0.89	0.89	0.88	423



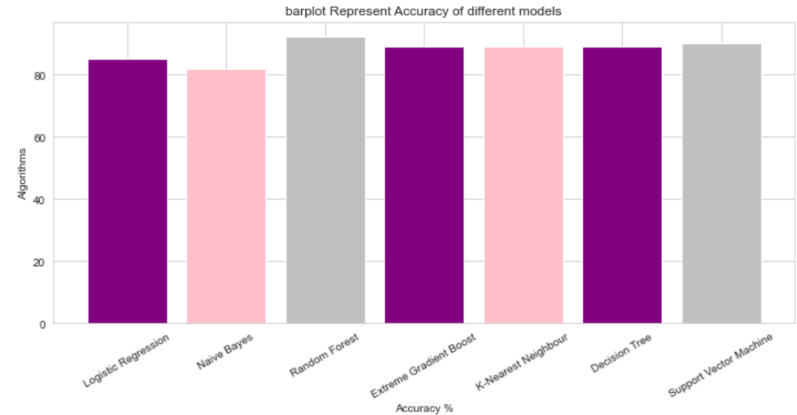
Support Vector Classifier

Accuracy of Support Vector Classifier: 90.0709219858156

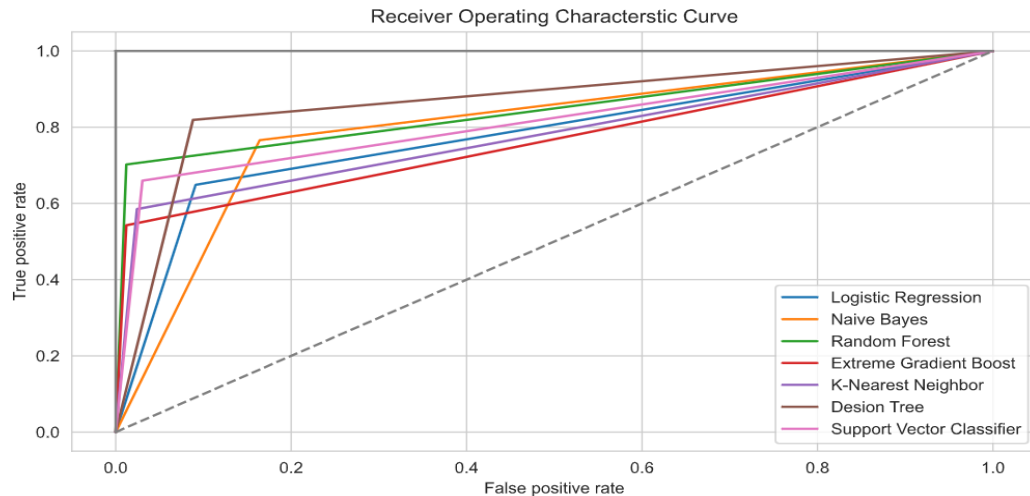
	precision	recall	f1-score	support
0	0.91	0.97	0.94	329
1	0.86	0.66	0.75	94
accuracy			0.90	423
macro avg	0.88	0.81	0.84	423
weighted avg	0.90	0.90	0.90	423



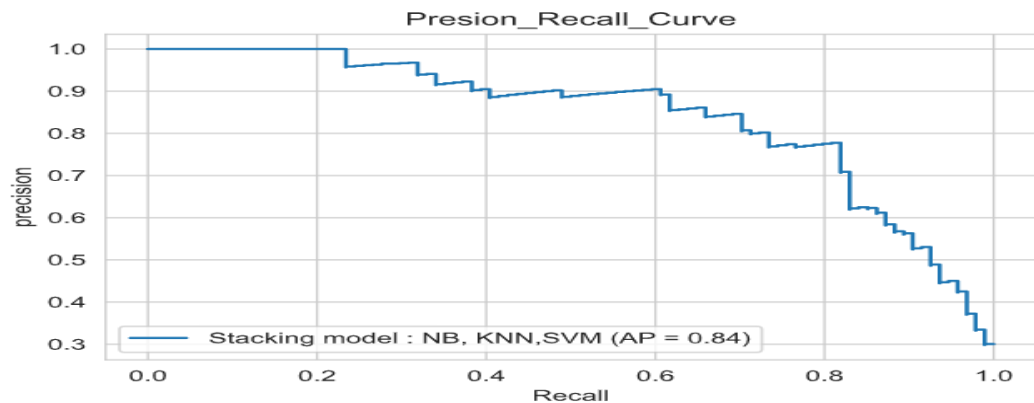
	Model	Accuracy
0	Logistic Regression	85.106383
1	Naive Bayes	82.033097
2	Random Forest	92.434988
3	Extreme Gradient Boost	88.888889
4	K-Nearest Neighbour	88.888889
5	Decision Tree	89.125296
6	Support Vector Machine	90.070922



Stacking



PRC for Stacking
Model



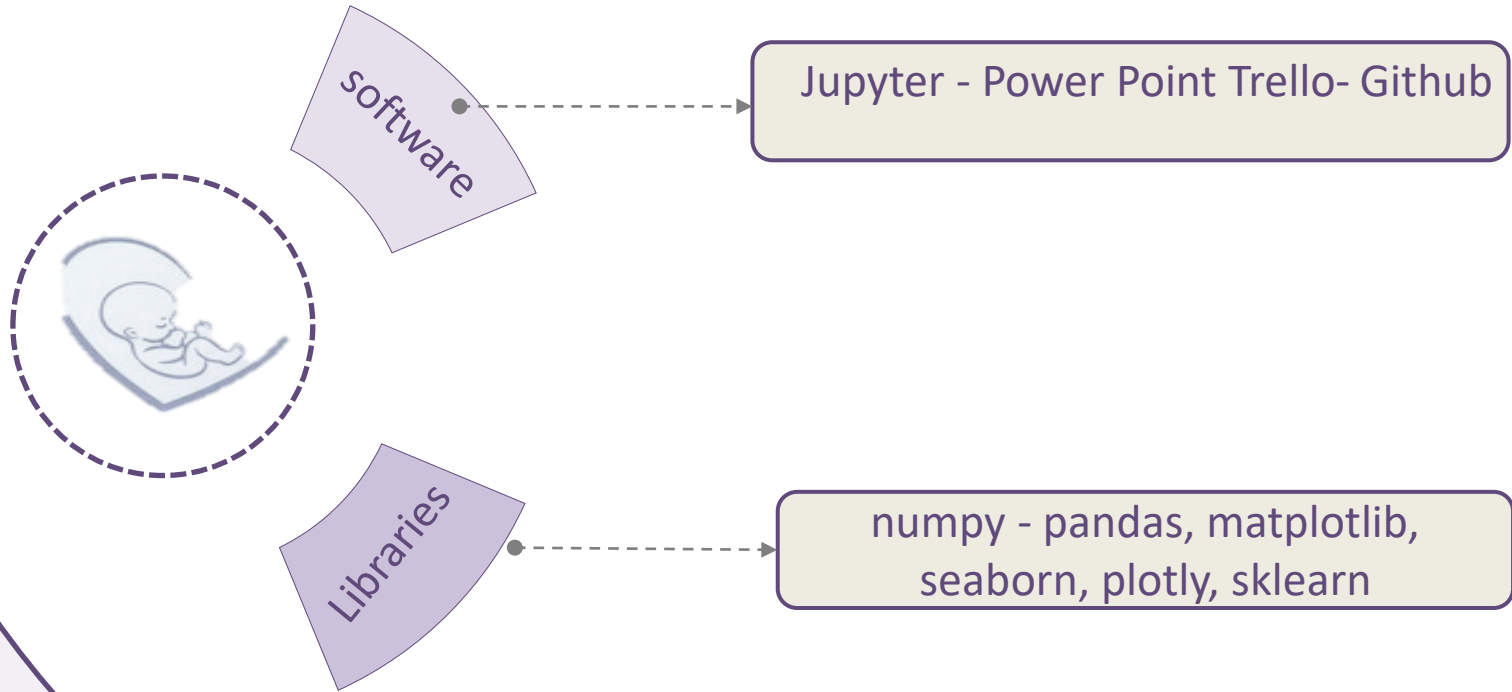
PRC for Models

Challanges



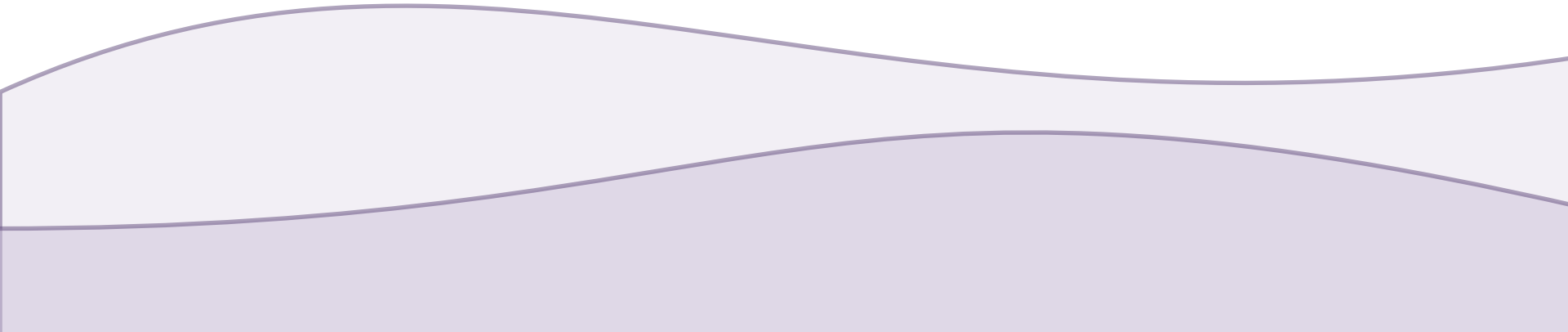
- 01 Handling Imbalance data
- 02 Find a suitable hyperparameter
- 03 Long execution time

Tools



○ Conclusion

Our identified key features are the same features that doctors look for during a visual analysis. Because our model is so precise, we can be confident that our model will predict the class at a better rate than the visual analysis, eliminating human error.



THANKS

