

FETALHEALTH CLASS MODELING

Presented by: Bayan Ali – Eman Albarakah – Yaqeen alhawaj



CONTANTS

Introduction Algorithm

Workflow Challenges

Data overview Tools

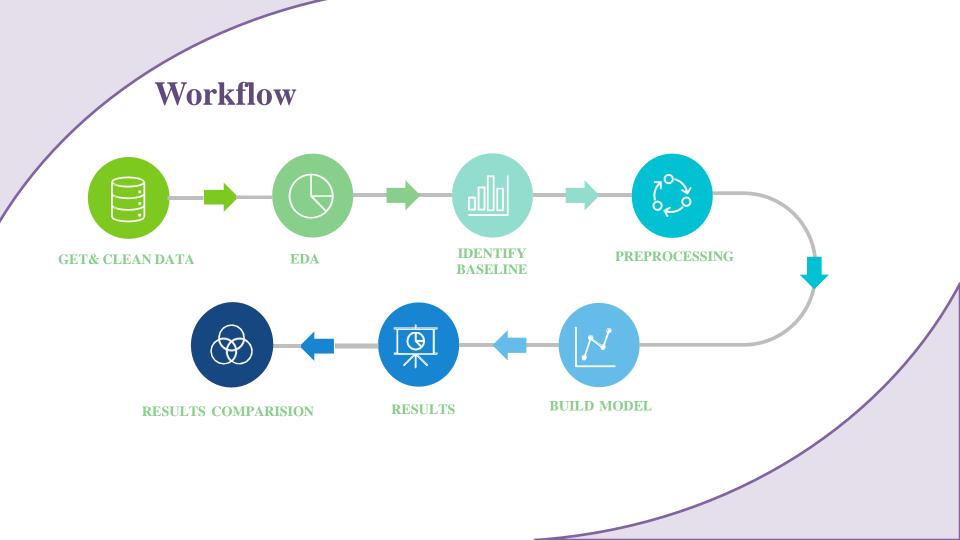
EDA Conclusion

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.



Data Overview

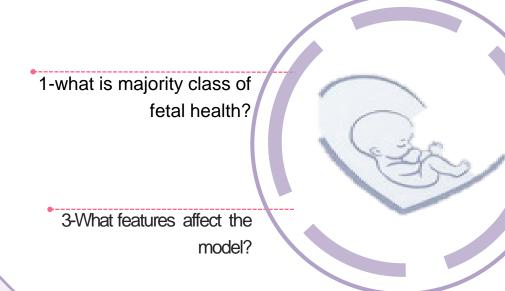
KaggleIt has 2126 rows and 22 columns

It has 2120 fows 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	Number of uterine contractions per second							
contractions								
light decelerations	Number of LDs per second							
severe	Number of SDs per second							
decelerations								
prolongueddecele	Number of PDs per second							
rations								
abnormalshortter	Percentage of time with abnormal short-term variability							
mvariability								

- **Exploring Data**
- Solve Outliers
- Duplicated



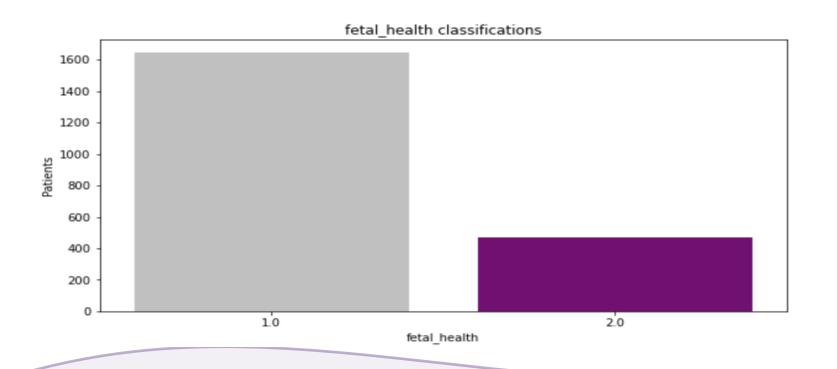
Business Questions



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?

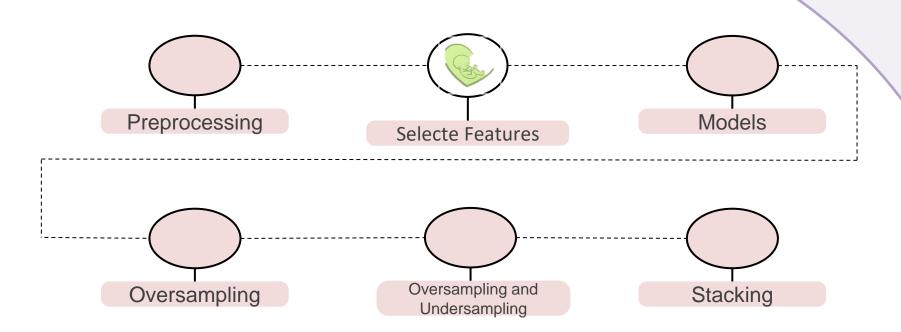
EDA Visualization



EDA Visualization

																							- 170	- 1.00
baseline_value	1				-0.16				-0.33								0.79		0.79	-0.18		0.25		1.00
accelerations	-0.082							-0.28		-0.39	-0.18		-0.16									-0.39		
fetal_movement	0.034												-0.15									5 0 0 6 1		- 0.75
Uterine_contractions	-0.15			1.		0 0064		-0.23		-0.22								-0.17				-0.26		
light_decelerations	-0.16					0.11				-0.28	-0.29		-0.56				-0.24	-0.46	-0.35			-0.032		
Severe_Dcelerations	-0.054					.1	0.012															880.0		- 0.50
Prolongued_Decelerations	-0.11					0.012	4	0.046	0.3	-0.16	-0.25		-0.28				-0.2	-0.36	-0.36		-0.22	0.34		
ABshort_variability	0.3	-0.28		-0.23	-0.12	0.034		- 1	-0.44	0.36	-0.31	-0.26	0.28	-0.16	-0.16	-0.18		0.15		-0.22		0.49		
mean_short_variability	0.33	0.25		0.32	0.53	0.052		-0.44	1	-0.51		0.63	-0.64	0.31	0.48	0.25	-0.26	-0.41	-0.34	0.59		-0.25		- 0.25
ercentage_long_variability	0.3	-0.39		-0.22	-0.28		-0.16	0.36	-0.51	1	-0.013	-0.38		-0.22	-0.25					-0.38		0.35		
mean_long_variability	-0.071	-0.18			-0.29	-0.04	-0.25	-0.31	-0.1		- 1	0.0087	-0 033	-0.045	-0.015	0.049				-0.16		-0.17		
histogram_width	-0.16	0.3		0.14	0.52	0.045	0.27	-0.26	0.63	-0.38		1	-0.9	0.65	0.74	0.32		-0.22	0.14	0.68		-0.14		- 0.00
histogram_min	0.36	-0.16	-0.15		-0.56		-0.28	0.28	-0.64	0.37		-0.9	- 1	-0.29	-0.67	-0.32				-0.63	-0.24	0.16		
histogram_max	0.31							-0.16		-0.22			-0.29	1	0.49	0.17	0.41	0.34	0.38			-0.051		
togram_number_of_peaks	-0.1							-0.16		-0.25		0.74	-0.67	0.49	12	0.29		-0.16				0.073		0.2
ogram_number_of_zeroes	0.071				0.26			-0.18	0.25			0.32	-0.32			1	-0.055	0.12	-0:076			0.056		
histogram_mode	0.79				-0.24		-0.2		-0.26						0.034							0.099		
histogram_mean	0.74			-0.17	-0.46		-0.36		-0.41			-0.22			-0.16		0.81		0.91	-0.26		0.056		0.5
histogram_median	0.79			-0.15	-0.35	-0.015	-0.36		-0.34	0.22		-0::1:4	0.38		-0.091	-0.076	0.9	0.91	1	-0.15		0.021		
histogram_variance	-0.18				0.59	0.029	0.27	-0.22	0.59	-0.38	-0.16	0.68	-0.63	0.42	0.53	0.23		-0.26	-0.15	1	0.069	-0.13		
histogram_tendency	0.29	0.031					-0.22	-0:01		0.064			-0.24								1	-0.07		0.7
fetal_health	0.25	-0.39	0.061	-0.26	-0.032	0.068	0.34	0.49	-0.25	0.35	-0.17	0.14	0.16	-0.051	-0.073	-0.056	0.099	0.056	0.021	-0.13	-0.07	1		
	e value	accelerations	Wement	ractions	erations	erations	erabons	anability	anability	anability	anability	m width	am min	m max	(_peaks	ZBroes	m_mode	mean_	medan	variance	andency	fital_health		
	hasein	acce	etal movi	lerine_contraction	lght_deceleration	Severe_Doeleration	gued_Decelera	4Bshort_variabilit	nean_short_variability	log v	en_long_vanability	histogram_width	histogram	histogram	umber_of_peak	number of zeroe	histogram	histogram	histogram_media	histogram_varianc	histogram_tendenci	輯		
				- E	20	8	olongue	25	mean	centage_long_variabilit	mean				gram_nu	am_ma	-	-	22	His	hsto			

Algorithm

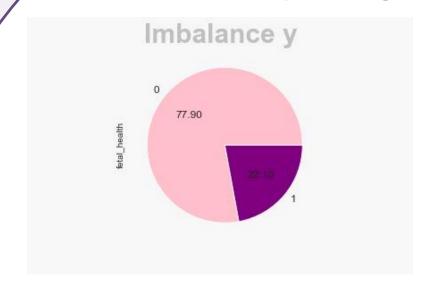


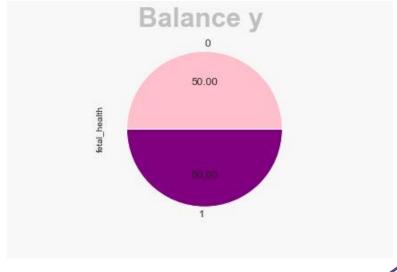
Selected Features

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

SMOTE

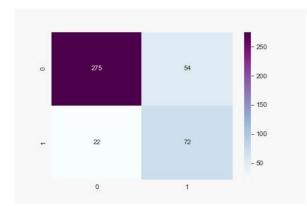
We found that we have a problem of imbalanced data set, namely that target of 1 is less than number of 0.

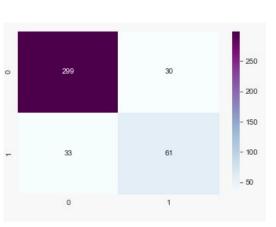




Naive Bayes

Logistic Regression





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
accura	су			0.82	423
macro a	IVg	0.75	0.80	0.77	423
weighted a	vg	0.85	0.82	0.83	423

Accuracy of Logistic Regression: 85.1063829787234

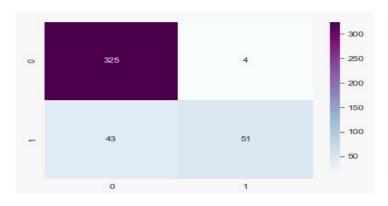
	precision	recall	f1-score	support
0	0.90	0.91	0.90	329
1	0.67	0.65	0.66	94
accuracy			0.85	423
macro avg	0.79	0.78	0.78	423
weighted avg	0.85	0.85	0.85	423

Random Forest



Accuracy	of R	andom Forest:	92.4349	8817966903	
		precision	recall	f1-score	support
	0	0.92	0.99	0.95	329
	1	0.94	0.70	0.80	94
accur	racy			0.92	423
macro	avg	0.93	0.84	0.88	423
weighted	avg	0.93	0.92	0.92	423

Extreme Gradient Boost



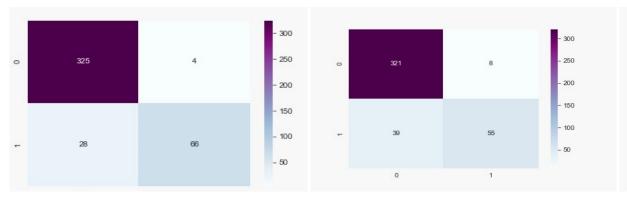
Accuracy	of E	xtreme Gradı	ent Boost	: 88.88888	388888889
		precision	recall	f1-score	support
	0	0.88	0.99	0.93	329
	1	0.93	0.54	0.68	94
accur	асу			0.89	423
macro	avg	0.91	0.77	0.81	423
weighted	avg	0.89	0.89	0.88	423

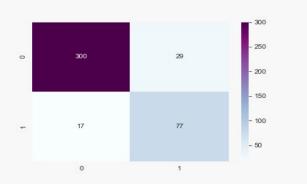
Decision Tree

K-Nearest Neighbor

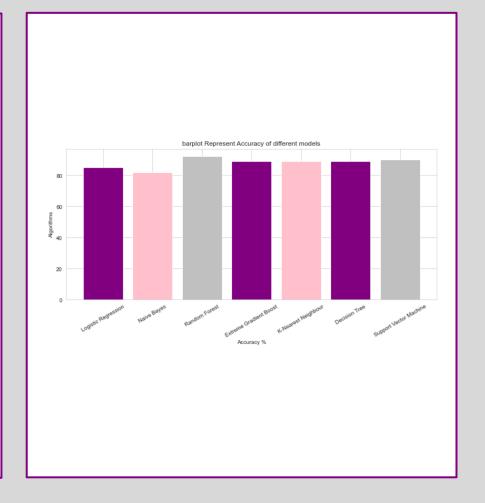
Support Vector Classifier

Accuracy of	DecisionTree(Classifier	: 89.12529	550827422	Accuracy of k	(-NeighborsCl	assifier:	88.888888	88888889	Accuracy of s	Support Vecto	r Classif	ier: 90.07	092 <mark>1</mark> 9858156
	precision	recall	f1-score	support		precision	recall	f1-score	support		precision	recall	f1-score	support
0	0.95	0.91	0.93	329	0	0.89	0.98	0.93	329	0	0.91	0.97	0.94	329
1	0.73	0.82	0.77	94	1	0.87	0.59	0.70	94	1	0.86	0.66	0.75	94
accuracy			0.89	423				0.00	422	accuracy			0.90	423
macro avg		0.87	0.85	423	accuracy			0.89	423	macro avg	0.88	0.81	0.84	423
U					macro avg	0.88	0.78	0.82	423	weighted avg	0.90	0.90	0.90	423
weighted avg	0.90	0.89	0.89	423	weighted avg	0.89	0.89	0.88	423					

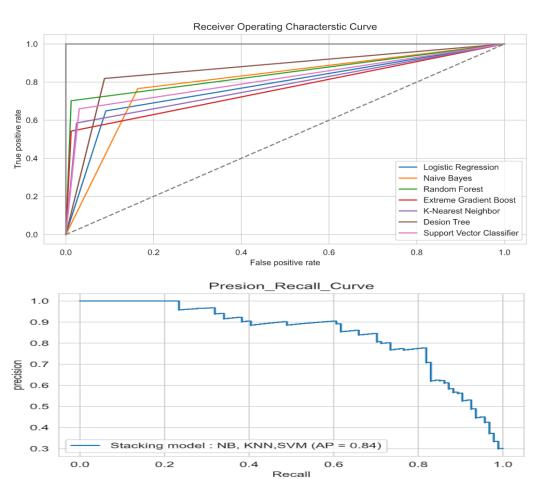




	Model	Accuracy
0	Logistic Regression	85.106383
1	Naive Bayes	82.033097
2	Random Forest	92.434988
3	Extreme Gradient Boost	88.88889
4	K-Nearest Neighbour	88.88889
5	Decision Tree	89.125296
6	Support Vector Machine	90.070922



Stacking



PRC for Stacking Model

PRC for Models

Challanges



O1 Handling Imbalance data

Find a suitable hyperparameter 02

Long execution time

Tools Jupyter - Power Point Trello- Github numpy - pandas, matplotlib, seaborn, plotly, sklearn

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.

