

Instructions and Results

October 15, 2021

1 Preprocessing

Preprocessing code are store in preprocessing/. The set of preselected features are listed in ./pre-processing/preeclampsia_features.csv

Runing below code will generate dataset for maternal visit 1, 2, 3 and 4 in the ./data/ folder

Size of V1 dataset (1665 Women, 63 Features)

Size of V2 dataset (1665 Women, 103 Features)

Size of V3 dataset (1665 Women, 135 Features)

Size of V2 dataset (1665 Women, 182 Features)

```
[ ]: ! python preprocessing/generate_dataset.py
```

2 Running Experiments

We are running a 5 fold cross vaildation 20 time which result in 100 trails for the XGBoost model.

The main.py runs the test trail for V1, V2, V3, and V4. Running the below code will generate the results and save them as pickle file in the ./result/{Visit #}/{Classification Model}/ folder. It will also produce the feature importances and the AUC and store in in the ./result/summary/ folder. Sensitivity, specificity, PPV and NPV will also be outputed.

```
[ ]: !python main.py
```

This is a summarization of the results from our experiments

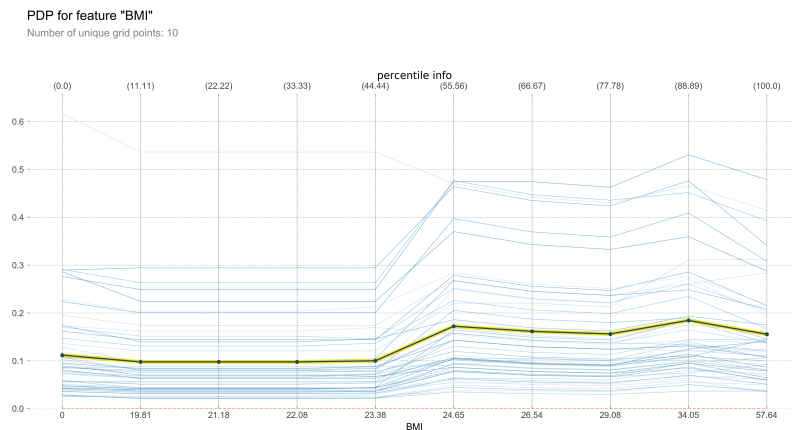
	AUC	Sensitivity (TP/ TP+FN)	Specificity (TN/ FP + TN)	PPV (TP/ TP +FP)	NPV (TN/FN+TN)
V1	0.66	0.61	0.61	0.61	0.61
V2	0.70	0.62	0.67	0.66	0.64
V3	0.73	0.65	0.69	0.68	0.66
V4	0.80	0.70	0.74	0.73	0.71
Target	257				
Non-Target	1408				

model_result

3 PDP Plots

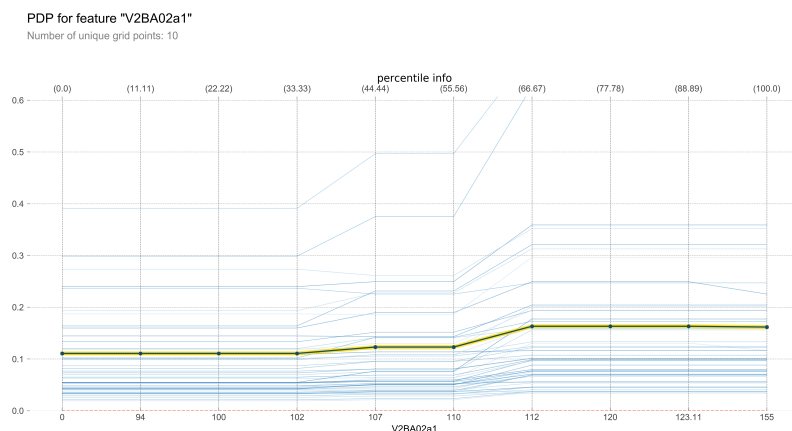
Code for generating the PDP Plots can be located on
pdps.ipynb.

A well-known risk factor for PE is having an elevated BMI, specifically over 30 km/m2. Looking at the PDP plot for this feature, we observe a significant risk increase in PE at around 22.41 km/m2.

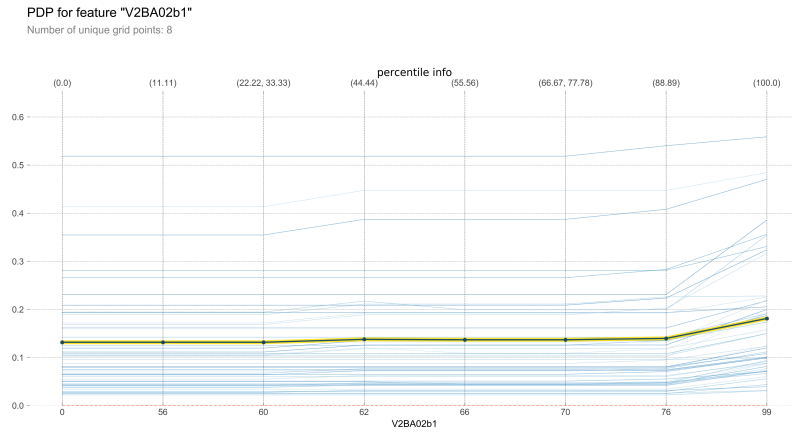


BMI

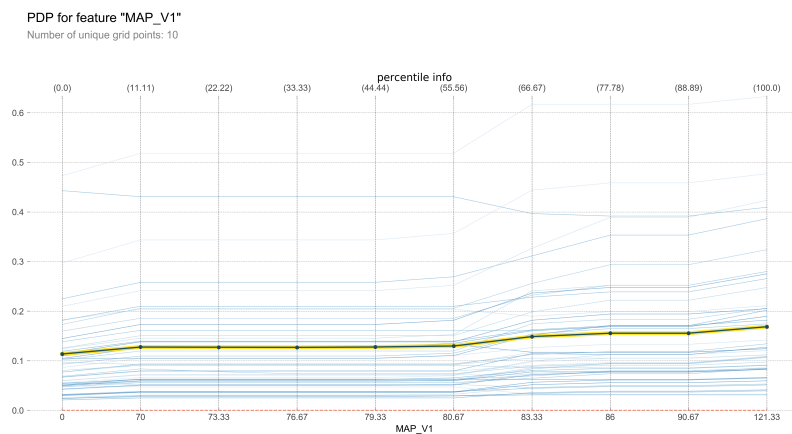
The classic risk factor for PE is an elevated BP. Looking at the PDP plot for the relevant features at Visit 1, we witnessed a significant increase in the risk of PE with a systolic reading of 110 mmHg or higher, and by Visit 2 this number drops to 102 mmHg. The diastolic reading did not exhibit such a pronounced increase in the risk of PE, but we did observe a slight increase above 78 mmHg. Looking at the Mean Arterial Pressure (MAP) at Visit 1, we see an increase in risk at 80.67 mmHg.



(V2B) Resting blood pressure - Systolic measure 1 (mmHg)



(V2B) Resting blood pressure - Diastolic measure 1 (mmHg)



Mean Arterial Pressure

4 Running Fairness Checks on Model

Fairness checks code are in util/fairness.py running. It will plot the values of the fairness metrics of the 100 trail conducted perviously by look at the stored files in the ./result/ folder

Run the below command to get the fairness check plot and Ceteris Paribus Cutoff plot

```
[ ]: !python util/fairness.py
```

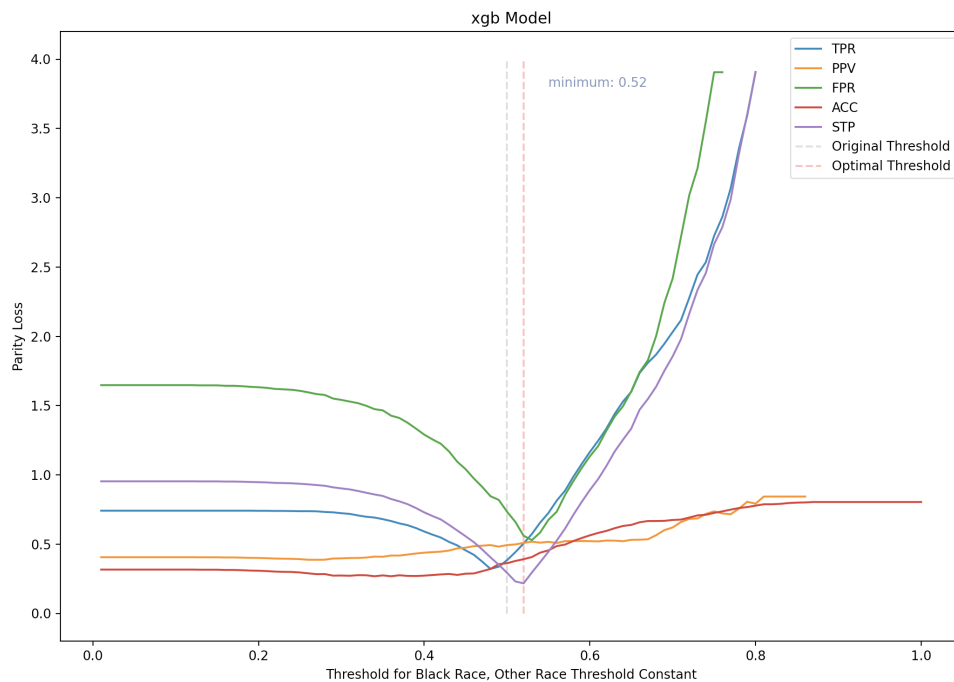
5 Fairness check of Original Model

Predictive equality ratio were high for American American Race. Results of an over diagnosis.



Fairness_check_original

6 Ceteris Paribus Cutoff plot



Ceteris_Paribus_Cutoff

7 Fairness Plot After Selecting Optimal Threshold



Fairness_check_optimal

8 Appendix

For a full list of Top features please refer to ./supplements/ folder

9 Top 20 Features from V1

Feature	Feature Importance	Description
V1BA06a1	0.0520	(V1B) Resting blood pressure - Systolic (mm Hg) measure 1
BMI	0.0478	BMI (kg/m ²) at visit 1 calculated from height and weight (form V1B)
V1BA04a	0.0461	(V1B) Waist circumference over iliac crest - Measure 1 (cm)
MAP_V1	0.0411	(V1 Systolic BP + 2 * V1 Diastolic BP)/3
V1BA07a	0.0401	(V1B) Neck circumference - Measure 1 (cm)
ENDOGUN	0.0290	Endoglin (ng/mL)
CLAC01f2	0.0269	(CLA) Inhibin A (DIA) Results - Multiples of the median (MoM)
DT_PFAT	0.0267	Polyunsaturated fatty acids, grams
VEGF	0.0248	Vascular endothelial growth factor (VEGF) (pg/mL)
AFP	0.0233	Alpha fetoprotein (AFP) (IU/mL)
ADAM12	0.0227	ADAM-12 (ng/mL)
DT_SFAT	0.0226	Saturated fat, grams
DT_CHOL	0.0221	Cholesterol, mg
VXXB01aa_V1a	0.0202	(VXX) Medical conditions or diagnoses, High blood pressure (hypertension) - Condition noted, Visit 1
VXXB01ar_V1a	0.0201	(VXX) Medical conditions or diagnoses, History of blood clots (thrombosis or thromboembolism) or stroke - Condition noted, Visit 1
U1CD07	0.0198	(U1C) Right Uterine Artery - Pulsatility Index (PI)
DT_MFAT	0.0195	Monounsaturated fatty acids, grams
Age_at_V1	0.0195	Age (years) at visit 1 calculated from DOB (V1AB01) and the form completion date
CLAB02d1b	0.0191	(CLA) First trimester serum screen PAPP-A Results - μ U/mL
fbHCG	0.0188	Free beta human chorionic gonadotropin (fbHCG) (ng/mL)

Top 20 Feature V1

10 Top 20 Features from V2

Feature	Feature Importance	Description
BMI	0.0259	BMI (kg/m ²) at visit 1 calculated from height and weight (form V1B)
VXXB01ar_V2a	0.0246	(VXX) Medical conditions or diagnoses, History of blood clots (thrombosis or thromboembolism) or stroke - Condition noted, Visit 2
MAP_V1	0.0231	(V1 Systolic BP + 2 * V1 Diastolic BP)/3
V1BA04a	0.0219	(V1B) Waist circumference over iliac crest - Measure 1 (cm)
V1BA07a	0.0210	(V1B) Neck circumference - Measure 1 (cm)
U2CD09	0.0209	(U2C) Right Uterine Artery - Early diastolic notch present
V2BA02a1	0.0196	(V2B) Resting blood pressure - Systolic measure 1 (mmHg)
MAP_V2	0.0187	(V2 Systolic BP + 2 * V2 Diastolic BP)/3
V1BA06a1	0.0171	(V1B) Resting blood pressure - Systolic (mm Hg) measure 1
ENDOGLIN_1	0.0164	Endoglin (ng/mL)
ENDOGLIN	0.0162	Endoglin (ng/mL)
U2CC09	0.0161	(U2C) Left Uterine Artery - Early diastolic notch present
VXXB01ah_V2a	0.0155	(VXX) Medical conditions or diagnoses, Valvular heart disease - Condition noted, Visit 2
U2CC05	0.0151	(U2C) Left Uterine Artery - Systolic/Diastolic Ratio (S/D)
VXXB01ap_V2a	0.0146	(VXX) Medical conditions or diagnoses, Thrombocytopenia - Condition noted, Visit 2
INHIBINA_1	0.0142	Inhibin A (pg/mL)
VXXB01aa_V2a	0.0140	(VXX) Medical conditions or diagnoses, High blood pressure (hypertension) - Condition noted, Visit 2
U2CC07	0.0139	(U2C) Left Uterine Artery - Pulsatility Index (PI)
PLGF_1	0.0136	Placental growth factor (PLGF) (pg/mL)
VXXB01ak_V2a	0.0134	(VXX) Medical conditions or diagnoses, Cardiac arrhythmias - Condition noted, Visit 2

Top 20 Feature V2

11 Top 20 Features from V3

Feature	Feature Importance	Description
V3BA02a1	0.0446	(V3B) Resting blood pressure - Systolic measure 1 (mmHg)
MAP_V3	0.0445	(V3 Systolic BP + 2 * V3 Diastolic BP)/3
BMI	0.0204	BMI (kg/m ²) at visit 1 calculated from height and weight (form V1B)
U2CD09	0.0166	(U2C) Right Uterine Artery - Early diastolic notch present
VXXB01az_V2a	0.0145	(VXX) Medical conditions or diagnoses, Liver/gall bladder disease - Recent use of medication, Visit 2
ENDOGLIN_1	0.0140	Endoglin (ng/mL)
V1BA06a1	0.0133	(V1B) Resting blood pressure - Systolic (mm Hg) measure 1
ENDOGLIN	0.0133	Endoglin (ng/mL)
U2CC09	0.0132	(U2C) Left Uterine Artery - Early diastolic notch present
U3CC06	0.0129	(U3C) Left Uterine Artery - Resistance Index (RI)
V2BA02a1	0.0125	(V2B) Resting blood pressure - Systolic measure 1 (mmHg)
MAP_V1	0.0124	(V1 Systolic BP + 2 * V1 Diastolic BP)/3
VXXB01aa_V2a	0.0122	(VXX) Medical conditions or diagnoses, High blood pressure (hypertension) - Condition noted, Visit 2
DT_CHOL	0.0122	Cholesterol, mg
V1BA07a	0.0120	(V1B) Neck circumference - Measure 1 (cm)
U2CC07	0.0119	(U2C) Left Uterine Artery - Pulsatility Index (PI)
CLAD01a2	0.0112	(CLA) Third Trimester Lab Studies (> 24 0 weeks) Blood count, Lowest hemoglobin during third trimester before labor and delivery - Result g/dL
INHIBINA_1	0.0107	Inhibin A (pg/mL)
U2CC05	0.0106	(U2C) Left Uterine Artery - Systolic/Diastolic Ratio (S/D)
U3AB04	0.0105	(U3A) Fetal cardiac activity - Abdominal Circumference (cm)

Top 20 Feature V3

12 Top 20 Features from V4

Feature	Feature Importance	Description
V3BA02a1	0.0439	(V3B) Resting blood pressure - Systolic measure 1 (mmHg)
VXXB01aa_FA	0.0409	(VXX) Medical conditions or diagnoses, High blood pressure (hypertension) - Final Assessment: Condition present
MAP_V3	0.0278	(V3 Systolic BP + 2 * V3 Diastolic BP)/3
V3BA02b1	0.0209	(V3B) Resting blood pressure - Diastolic measure 1 (mmHg)
VXXB01aa_V2a	0.0198	(VXX) Medical conditions or diagnoses, High blood pressure (hypertension) - Condition noted, Visit 2
CMAH01b3	0.0171	(CMA) Results of complete blood count: - Hemoglobin (g/dL)
BMI	0.0162	BMI (kg/m^2) at visit 1 calculated from height and weight (form V1B)
VXXB01aa_V1a	0.0150	(VXX) Medical conditions or diagnoses, High blood pressure (hypertension) - Condition noted, Visit 1
V1BA06a1	0.0147	(V1B) Resting blood pressure - Systolic (mm Hg) measure 1
CLAE08a1	0.0122	(CLA) First formal proteinuria quantification (if done prior to 20/0 weeks project EGA) - Total protein (mg/24hr) (a)
V1BA07a	0.0118	(V1B) Neck circumference - Measure 1 (cm)
VXXB01av_V2a	0.0116	(VXX) Medical conditions or diagnoses, Rheumatoid arthritis - Recent use of medication, Visit 2
U3CC06	0.0115	(U3C) Left Uterine Artery - Resistance Index (RI)
VXXB01aa_V4a	0.0113	(VXX) Medical conditions or diagnoses, High blood pressure (hypertension) - Condition noted, Visit 4
V2BA02a1	0.0108	(V2B) Resting blood pressure - Systolic measure 1 (mmHg)
U2CD09	0.0108	(U2C) Right Uterine Artery - Early diastolic notch present
V1BA04a	0.0107	(V1B) Waist circumference over iliac crest - Measure 1 (cm)
VXXB01az_V2a	0.0106	(VXX) Medical conditions or diagnoses, Liver/gall bladder disease - Recent use of medication, Visit 2
CMAH01b6	0.0100	(CMA) Results of complete blood count: - Platelet count x10^3 / mm^3 (x10^3 /μL)
ENDOGLIN_1	0.0098	Endoglin (ng/mL)
DT_MFAT	0.0097	Monounsaturated fatty acids, grams

Top 20 Feature V4

[]: