CounterFactuals using CDC Data RESULTS

WORKFLOW :

1) We have 5.0 GB of Birth Data File taken form “Centres for disease control and prevention (CDC)” datasets.
2) https://www.cdc.gov/nchs/data_access/vitalstatsonline.htm
3) Divide into 15 subfiles contains strings of 330.1 MB.
4) Preprocess these 15 sub-files to .csv files.
5) Combined all these 15 files. Each of (38744×214).
6) Chosen the useful columns from from these 214 columns.
7) 71 Inputs columns,
8) Found that 62 columns which are useful.
9) 61 Input Columns, 1 Output column as “Condition”

   Condition= ”0” means “No Abnormal Conditions”

   Condition= ”1” means “Abnormal Conditions”.

INPUT DATA :

In [4]:
```
data5
```

Out[4]:

	Birth Place	Mother's Single Years of Age	Mother's Nativity	Residence Status	Mother's Hispanic Origin	Mother's Race/Hispanic Origin	Mother's Education	Father's Combined Age	Father's Hispanic Origin	Father's Race/Hispanic Origin	Father's Education	Prior Other Terminations	Month Prenatal Care Began	N Pr
0	1	42	1	1	0.0	1.0	1.0	90.0	0.0	6.0	1.0	1.0	21.0	
1	1	26	1	1	5.0	1.0	1.0	80.0	0.0	5.0	1.0	1.0	91.0	
2	1	24	2	1	4.0	1.0	1.0	50.0	1.0	3.0	1.0	2.0	11.0	
3	1	21	2	1	4.0	1.0	1.0	0.0	1.0	3.0	1.0	1.0	31.0	
4	1	30	2	1	4.0	1.0	1.0	50.0	4.0	3.0	1.0	3.0	91.0	
...	...	...	...	...	...	...	...	...	...	...	...	...	...	
165516	1	27	1	2	5.0	1.0	1.0	60.0	5.0	3.0	1.0	4.0	31.0	
165517	1	34	2	1	4.0	1.0	1.0	30.0	4.0	8.0	1.0	2.0	21.0	
165518	1	20	1	1	9.0	1.0	1.0	90.0	9.0	9.0	1.0	1.0	91.0	
165519	1	19	2	1	4.0	1.0	1.0	0.0	4.0	3.0	1.0	1.0	31.0	
165520	1	29	1	1	1.0	1.0	4.0	7.0	1.0	1.0	0.0	11.0	14.0	

165521 rows × 62 columns

In [ ]:
```

```

Output Target (Conditions) :

```
In [4]: data5
```

Out[4]:

Down Syndrome	Suspected Chromosomal Disorder	Hypospadias	Infant Transferred	Infant Living at Time of Report	Infant Breastfed at Discharge	Successful External Cephalic Version	Cigarettes Before Pregnancy.1	Delivery Weight Recode	Five Minute APGAR Score	Combined Gestation – Detail in Weeks	Obstetric Estimate Edited	Birth Weight – Detail in Grams (Edited)	Conditions
N	N	N	N	Y	U	N	0.0	71.0	94.0	38.0	38.0	2500.0	1
N	N	N	N	Y	U	N	0.0	421.0	94.0	39.0	39.0	3020.0	1
N	N	N	N	Y	U	N	0.0	361.0	94.0	34.0	32.0	1925.0	1
N	N	N	N	Y	U	N	0.0	751.0	94.0	34.0	34.0	1970.0	0
N	N	N	N	Y	U	N	0.0	981.0	94.0	39.0	39.0	3850.0	1
...	...	...	...	...	...	...	...	...	...	...	...	...	...
N	N	N	N	Y	U	N	0.0	721.0	94.0	34.0	33.0	2470.0	1
N	N	N	N	Y	U	N	0.0	741.0	94.0	41.0	39.0	3020.0	1
N	N	N	N	Y	U	N	0.0	591.0	94.0	40.0	40.0	3665.0	1
N	N	N	N	Y	U	N	0.0	91.0	83.0	38.0	39.0	4026.0	1
N	N	N	N	N	U	1	0.0	219.0	51.0	18.0	18.0	255.0	1

```
In [ ]:
```

0 : No Abnormal Condition.
1 : Abnormal Condition.

CATEGORICAL COLUMNS :

```
In [62]: print(data5.columns)
```

```
Index(['Birth Plate', 'Mother's Single Years of Age', 'Mother's Nativity',
       'Residence Status', 'Mother's Hispanic Origin',
       'Mother's Race/Hispanic Origin', 'Mother's Education',
       'Father's Combined Age', 'Father's Hispanic Origin',
       'Father's Race/Hispanic Origin', 'Father's Education',
       'Prior Other Terminations', 'Month Prenatal Care Began',
       'Number of Prenatal Visits', 'Cigarettes Before Pregnancy',
       'Cigarette Recode', 'Mother's Height in Total Inches',
       'Body Mass Index', 'Weight Gain', 'Pre-pregnancy Diabetes',
       'Gestational Diabetes', 'Pre-pregnancy Hypertension',
       'Gestational Hypertension', 'Hypertension Eclampsia',
       'Infertility Treatment Used', 'Fertility Enhancing Drugs',
       'Asst. Reproductive Technology', 'Induction of Labor',
       'Augmentation of Labor', 'Steroids', 'Antibiotics', 'Chorioamnionitis',
       'Fetal Presentation at Delivery', 'Final Route & Method of Delivery',
       'Maternal Transfusion', 'Perineal Laceration', 'Ruptured Uterus',
       'Unplanned Hysterectomy', 'Admit to Intensive Care',
       'Assisted Ventilation (immediately)', 'Assisted Ventilation > 6 hrs',
       'Admission to NICU', 'Surfactant', 'Antibiotics for Newborn',
       'Seizures', 'Limb Reduction Defect', 'Cleft Lip w/ or w/o Cleft Palate',
       'Cleft Palate alone', 'Down Syndrome', 'Suspected Chromosomal Disorder',
       'Hypospadias', 'Infant Transferred', 'Infant Living at Time of Report',
       'Infant Breastfed at Discharge', 'Successful External Cephalic Version',
       'Cigarettes Before Pregnancy.1', 'Delivery Weight Recode',
       'Five Minute APGAR Score', 'Combined Gestation – Detail in Weeks',
       'Obstetric Estimate Edited', 'Birth Weight – Detail in Grams (Edited)',
       'Conditions'],
      dtype='object')
```

Categorical Columns are one-hot encoded for training.

NUMERIC COLUMNS :

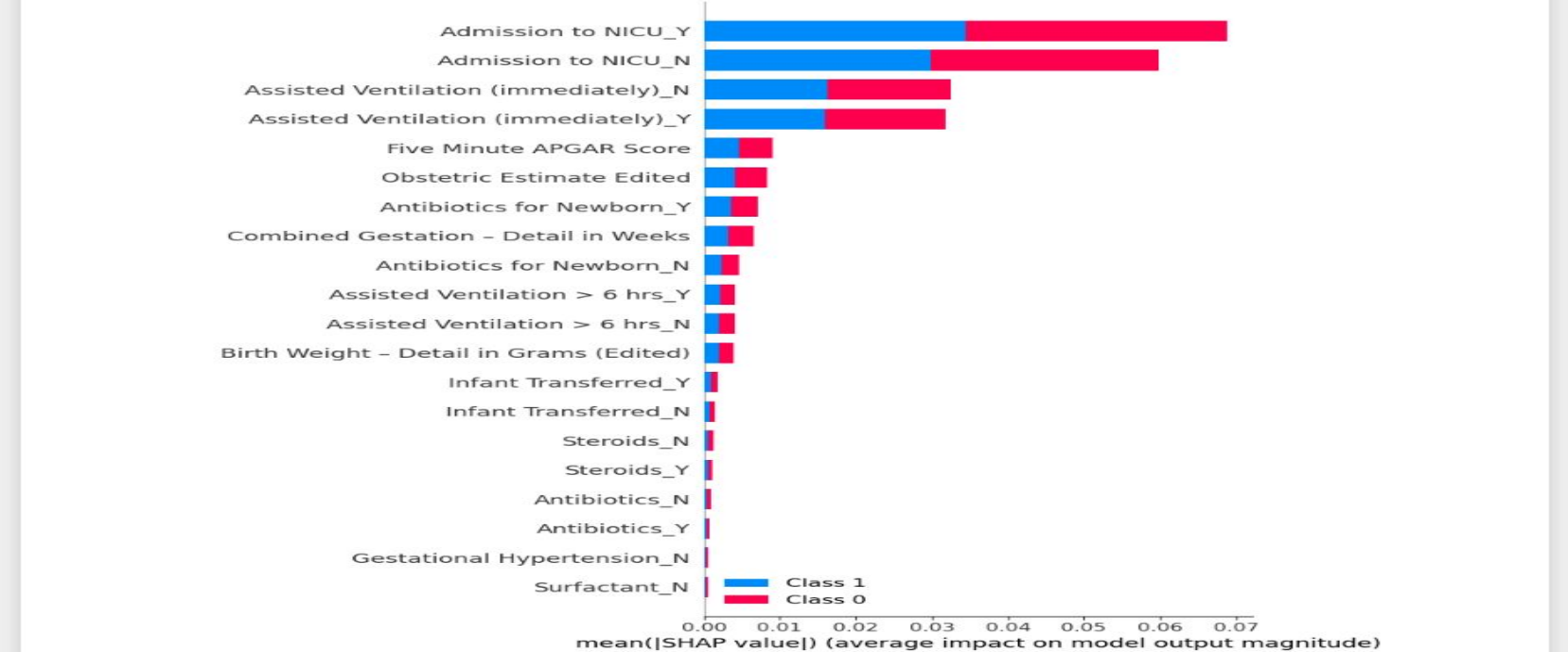
In [17]:

```
numeric_data=['Mother's Single Years of Age',
      'Father's Combined Age',
      'Prior Other Terminations',
      'Month Prenatal Care Began',
      'Number of Prenatal Visits',
      'Cigarettes Before Pregnancy',
      'Body Mass Index',
      'Weight Gain',
      'Mother's Height in Total Inches',
      'Delivery Weight Recode',
      'Five Minute APGAR Score',
      'Combined Gestation – Detail in Weeks',
      'Obstetric Estimate Edited',
      'Birth Weight – Detail in Grams (Edited)']
```

Numeric Values Changes during the CounterFactual Generations to Flip the class.

Detailed SHAP OUTPUT :

```
In [62]: # %% >> Visualize global features
         # Feature summary
         shap.summary_plot(shap_values, X_test)
```

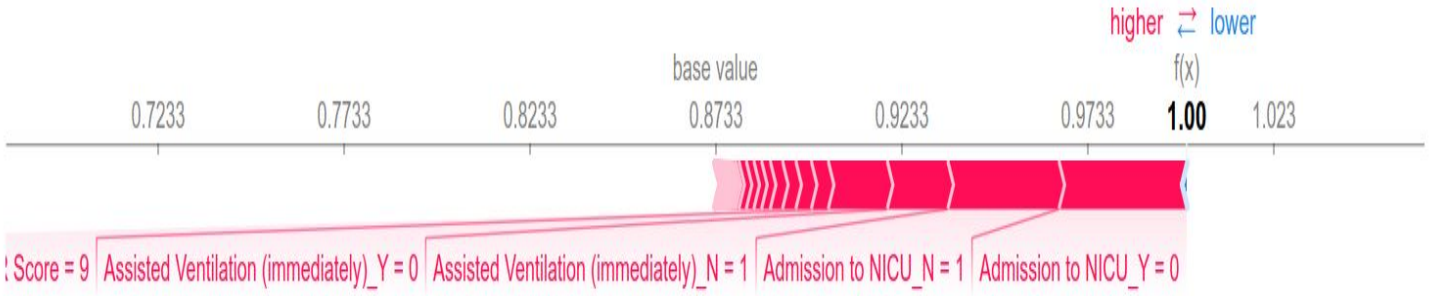


Top 5 Important Features contributing the Flipping the Class.
“Admission to NICU_Y”, “Assisted Ventilation”, “Five Minute APGAR Score”,”Anti-Biotics for NewBorn ”, “Combined Gestations”.

Force Plot for SHAP OUTPUT :

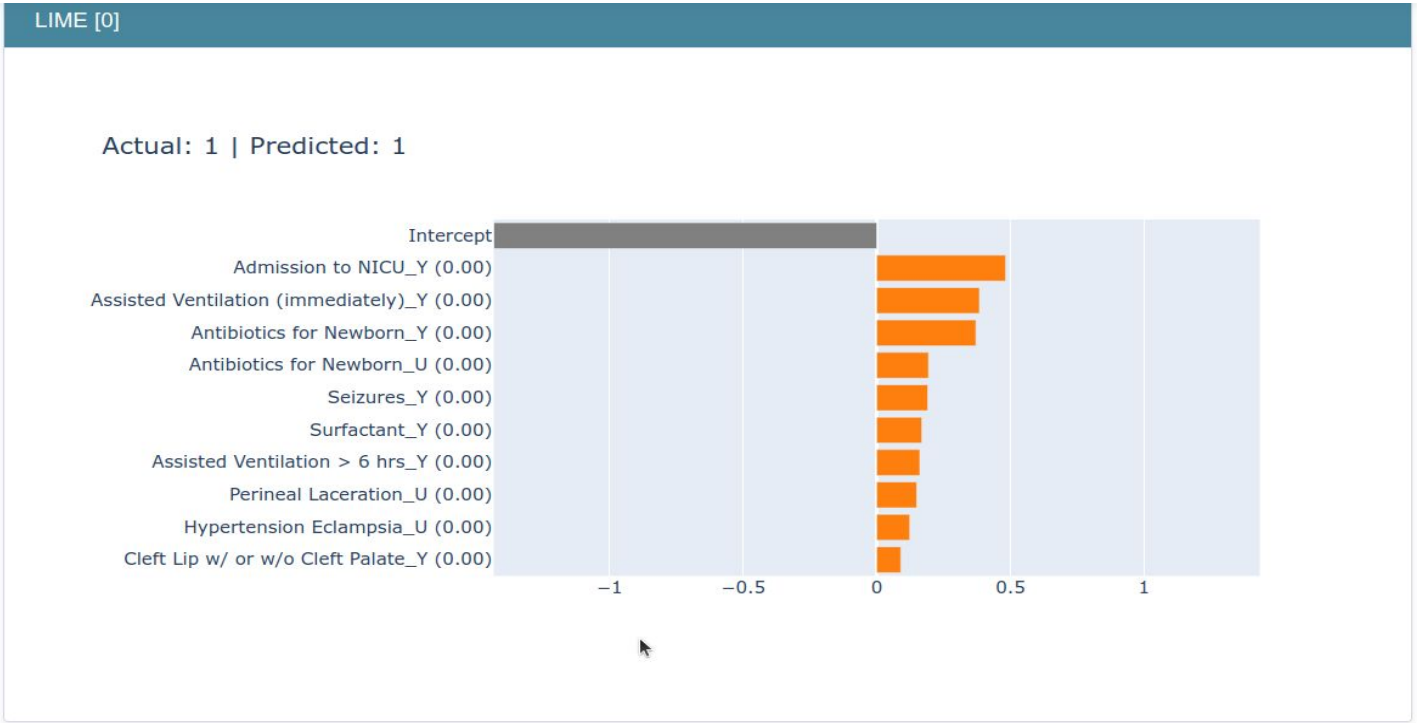
```
In [63]: shap.force_plot(explainer.expected_value[1],
                         shap_values[1],
                         X_test[start_index:end_index]) # for values
```

Out[63]:



Top 5 Important Features contributing the Flipping the Class.
“Admission to NICU_Y”, Probability : 0.9733
“Assisted Ventilation”, Probability : 0.9233

LIME OUTPUT :



“Admission to NICU_Y”, “Assisted Ventilation (immediately)”, Antibiotics for Newborn_Y are most contributing/important features as per “Local interpretation”.

COUNTERFACTUAL GENERATIONS :

1) 1 —> 0 CLASS CHANGE

In [73]:

```
input_datapoint = X_test[38:39]
cf = explainer.generate_counterfactuals(input_datapoint,
                                        total_CFs=5,
                                        desired_class="opposite")
cf.visualize_as_dataframe(show_only_changes=True)
```

```
100%|██████████████████████████████████████████████| 1/1 [00:00<00:00,  1.48it/s]
Query instance (original outcome : 1)
```

	Father's Combined Age	Prior Other Terminations	Month Prenatal Care Began	Number of Prenatal Visits	Cigarettes Before Pregnancy	Mother's Height in Total Inches	Body Mass Index	Weight Gain	Cigarettes Before Pregnancy.1	Delivery Weight Recode	Five Minute APGAR Score	Combined Gestation – Detail in Weeks	Obstetric Estimate Edited	Birth Weight – Detail in Grams (Edited)	Conditions
	99.0	0.0	0.0	0.0	0.0	66.0	15.8	10.0	0.0	108.0	9.0	41.0	41.0	4090.0	1

```
Diverse Counterfactual set (new outcome: 0)
```

	Father's Combined Age	Prior Other Terminations	Month Prenatal Care Began	Number of Prenatal Visits	Cigarettes Before Pregnancy	Mother's Height in Total Inches	Body Mass Index	Weight Gain	Cigarettes Before Pregnancy.1	Delivery Weight Recode	Five Minute APGAR Score	Combined Gestation – Detail in Weeks	Obstetric Estimate Edited	Birth Weight – Detail in Grams (Edited)	Conditions
	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.0
	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.0
	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.0
	-	-	-	-	-	-	-	27.3	-	-	-	-	-	-	0.0
	-	-	-	71.3	-	-	-	-	-	-	-	-	-	-	0.0

Class changes : “0” : Abnormal Conditions ,

“1” : No Abnormal Conditions.

1→0 FEATURES CHANGE :

```
In [73]:
input_datapoint = X_test[38:39]
cf = explainer.generate_counterfactuals(input_datapoint,
                                        total_CFs=5,
                                        desired_class="opposite")
cf.visualize_as_dataframe(show_only_changes=True)
```

```
100%|████████████████████████████████████████| 1/1 [00:00<00:00,  1.48it/s]
Query instance (original outcome : 1)
```

Admit to ntensive Care_1	Admit to Intensive Care_N	Admit to Intensive Care_U	Admit to Intensive Care_Y	Assisted Ventilation (immediately)_N	Assisted Ventilation (immediately)_U	Assisted Ventilation (immediately)_Y	Assisted Ventilation > 6 hrs_N	Assisted Ventilation > 6 hrs_U	Assisted Ventilation > 6 hrs_Y	Admission to NICU_N	Admission to NICU_U	Admissi to NICU
0	1	0	0	1	0	0	1	0	0	1	0	

```
Diverse Counterfactual set (new outcome: 0)
```

Admit to Intensive Care_1	Admit to Intensive Care_N	Admit to Intensive Care_U	Admit to Intensive Care_Y	Assisted Ventilation (immediately)_N	Assisted Ventilation (immediately)_U	Assisted Ventilation (immediately)_Y	Assisted Ventilation > 6 hrs_N	Assisted Ventilation > 6 hrs_U	Assisted Ventilation > 6 hrs_Y	Admission to NICU_N	Admission to NICU_U	Admis to NIC
-	-	-	-	-	-	-	-	-	-	0.0	-	
-	-	-	-	-	-	-	-	-	-	0.0	-	
-	-	-	-	-	-	-	-	-	-	-	-	
-	-	-	-	0.0	-	-	-	-	-	-	-	
-	-	-	-	0.0	-	-	-	-	-	-	-	

Features Change :

1) If there is Assisted Ventilation(Immediately) then “No Abnormal delivery”. (Pred=0)
2) If there is Admission to NICU is “YES” then “No Abnormal delivery”.(Pred=0)
3) “NO. of Prenatal Visits Increased from (0-> 71.3) ”,
4) “Weight Gain” increased from (10.0 -> 27.3).

COUNTERFACTUAL GENERATIONS :

1) 0 —> 1 CLASS CHANGE

```
In [50]:
input_datapoint = X_test[79:80]
cf = explainer.generate_counterfactuals(input_datapoint,
                                        total_CFs=5,
                                        desired_class="opposite")
cf.visualize_as_dataframe(show_only_changes=True)
```

```
100%|██████████| 1/1 [00:03<00:00,  3.27s/it]
Query instance (original outcome : 0)
```

Father's Combined Age	Prior Other Terminations	Month Prenatal Care Began	Number of Prenatal Visits	Cigarettes Before Pregnancy	Mother's Height in Total Inches	Body Mass Index	Weight Gain	Cigarettes Before Pregnancy.1	Delivery Weight Recode	Five Minute APGAR Score	Combined Gestation – Detail in Weeks	Obstetric Estimate Edited	Birth Weight – Detail in Grams (Edited)	Conditions
40.0	2.0	2.0	8.0	0.0	65.0	25.0	37.0	0.0	187.0	7.0	34.0	35.0	2370.0	0

```
Diverse Counterfactual set (new outcome: 1)
```

Prior Other Terminations	Month Prenatal Care Began	Number of Prenatal Visits	Cigarettes Before Pregnancy	Mother's Height in Total Inches	Body Mass Index	Weight Gain	Cigarettes Before Pregnancy.1	Delivery Weight Recode	Five Minute APGAR Score	Combined Gestation – Detail in Weeks	Obstetric Estimate Edited	Birth Weight – Detail in Grams (Edited)	Conditions
-	79.3	-	-	-	26.89999999999999	-	-	-	-	-	-	-	-
-	79.3	-	-	-	26.89999999999999	-	-	-	-	-	-	-	-
-	79.3	-	-	-	26.89999999999999	-	-	-	-	-	-	-	-
-	79.3	-	-	-	-	-	-	-	-	-	-	-	1.0
-	79.3	-	-	-	-	-	-	-	-	-	-	-	1.0

Class changes : Falling into Class “0” (No Abnormal Conditions) ,But we want to predict “1”, Abnormal Conditions.

0→1 FEATURES CHANGE :

```
In [50]:
input_datapoint = X_test[79:80]
cf = explainer.generate_counterfactuals(input_datapoint,
                                        total_CFs=5,
                                        desired_class="opposite")
cf.visualize_as_dataframe(show_only_changes=True)
```

```
100%|██████████| 1/1 [00:03<00:00,  3.27s/it]
Query instance (original outcome : 0)
```

	Hypospadias_U	Hypospadias_Y	Infant Transferred_N	Infant Transferred_U	Infant Transferred_Y	Infant Living at Time of Report_N	Infant Living at Time of Report_U	Infant Living at Time of Report_Y	Infant Breastfed at Discharge_N	Infant Breastfed at Discharge_U	Infant Breastfed at Discharge_Y	Succ E Ce Ver
	0	0	1	0	0	0	0	1	0	0	1	

```
Diverse Counterfactual set (new outcome: 1)
```

_N	Hypospadias_U	Hypospadias_Y	Infant Transferred_N	Infant Transferred_U	Infant Transferred_Y	Infant Living at Time of Report_N	Infant Living at Time of Report_U	Infant Living at Time of Report_Y	Infant Breastfed at Discharge_N	Infant Breastfed at Discharge_U	Infant Breastfed at Discharge_Y	Su V
-	-	-	-	1.0	-	-	-	-	-	-	-	
-	-	-	-	1.0	-	-	-	-	-	-	-	
-	-	-	-	1.0	-	-	-	-	-	-	-	
-	-	-	0.0	1.0	-	-	-	-	-	-	-	
-	-	-	0.0	1.0	-	-	1.0	-	-	-	-	

If below Features then Abnormal conditions :

1) Month Prenatal Care Began should not increased from 2.0 –>79.3,
2) Body Mass Index should not be increased > 26.8
3) If There is no Infant Transferred_U = 0.0.