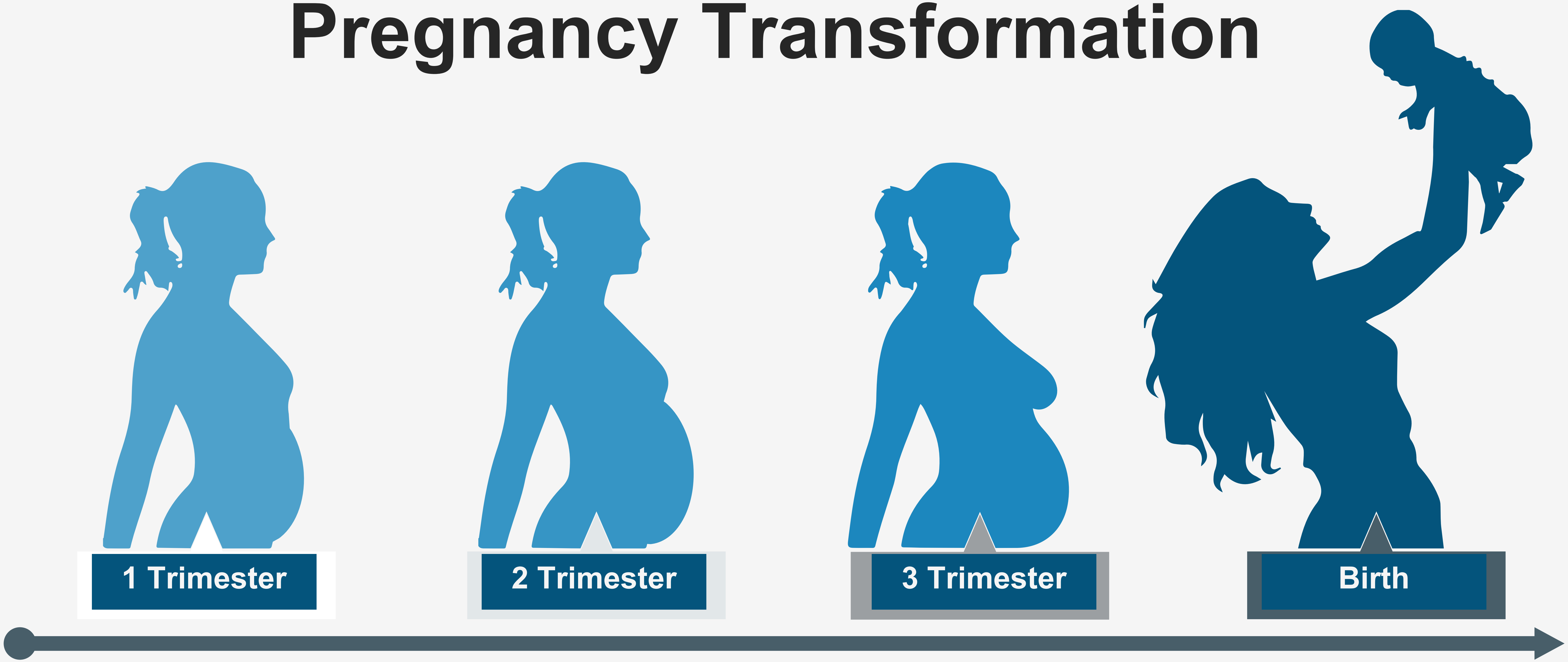
AI-based Potential Risks of Exposure to Air Pollution While Pregnant

Presented by:
Shakamuri Madhuri Vijaya Lakshmi - 20A21A05G1

Table of contents Review-2

Table of contents Review-2

Pregnancy Transformation



Introduction

- Air pollution, a combination of natural and artificial particles, impacts both indoors and outdoors. It can cause respiratory infections, reduced lung function, and asthma. Long-term health issues, such as lung cancer, stroke, heart attacks, and preterm delivery, are linked to air pollution. To reduce exposure, people should limit time in poor air quality areas and be aware of potential pollutants.

- Pregnant women and unborn children are at risk due to air pollution, requiring consultation with physicians and preventative measures like filter use and indoor confinement.

- Pregnancy-induced air pollution can lead to preterm labor, stillbirth, low birth weight, health concerns in the parent, lung development issues, and death after birth. Polluted areas increase the risk of these issues, with the risk highest during subsequent pregnancies. Exposure to pollutants can also disrupt a baby's development, increase the risk of and high blood pressure, and indirectly affect lung development.

Abstract

Air pollution, comprising ozone, particulate matter, nitrogen dioxide, sulfur dioxide, vehicle exhaust, building emissions, dust, and chemicals, poses diverse health risks including coughing, wheezing, eye irritation, respiratory diseases, fatigue, lung and heart damage, and cancer. Particularly vulnerable are pregnant women and their unborn children. This research investigates the impact of air pollution on pregnancy, highlighting potential adverse effects and proposing strategies to mitigate risks. Leveraging LightGBM and Shapley Additive explanations, the study elucidates significant features influencing air pollution and pregnancy outcomes, offering valuable insights for preventive measures.

Literature Survey

Reference [1] : Alexandra Grippo, Ajay A. published on 25-09-2022.

Air Pollution exposure during pregnancy and spontaneous abortions and stillbirth

- A total of 35 studies that are included in this review. Seventeen studies focused on spontaneous abortion, four of which focused on occupational exposures and spontaneous abortion, and 22 studies focused on stillbirth. Four studies investigated both spontaneous abortion and stillbirth as the outcome.

- The studies included in this review varied by population, geographic location, study design and exposure assessment. Study designs included were ecological, time series, cross-sectional, case-control and cohort study.

Literature Survey

Reference [2] :Prafulla Shriyan, Deepa Ravi, Giridhara R Babu, Yamuna A
published on 19-10-2021.

Ambient and Indoor Air Pollution in Pregnancy and the risk of Low birth weight and Ensuing Effects in Infants: A Cohort Study in Bangalore, South India [version 1; referees...

- A prospective cohort study will be conducted on 516 pregnant women in Bangalore's urban slums to assess air pollution levels, including PM10, PM2.5, and CO. The study will follow the children until they are two years old, evaluating the association between pollutants and low birth weight (LBW).

- PM10 and PM2.5 are particulate matter particles with varying sizes, with PM10 referring to 10 micrometers or less and PM2.5 referring to 2.5 micrometers or less. Both can cause health issues, with PM2.5 being more harmful due to deeper lung penetration.

EXISTING SYSTEMS

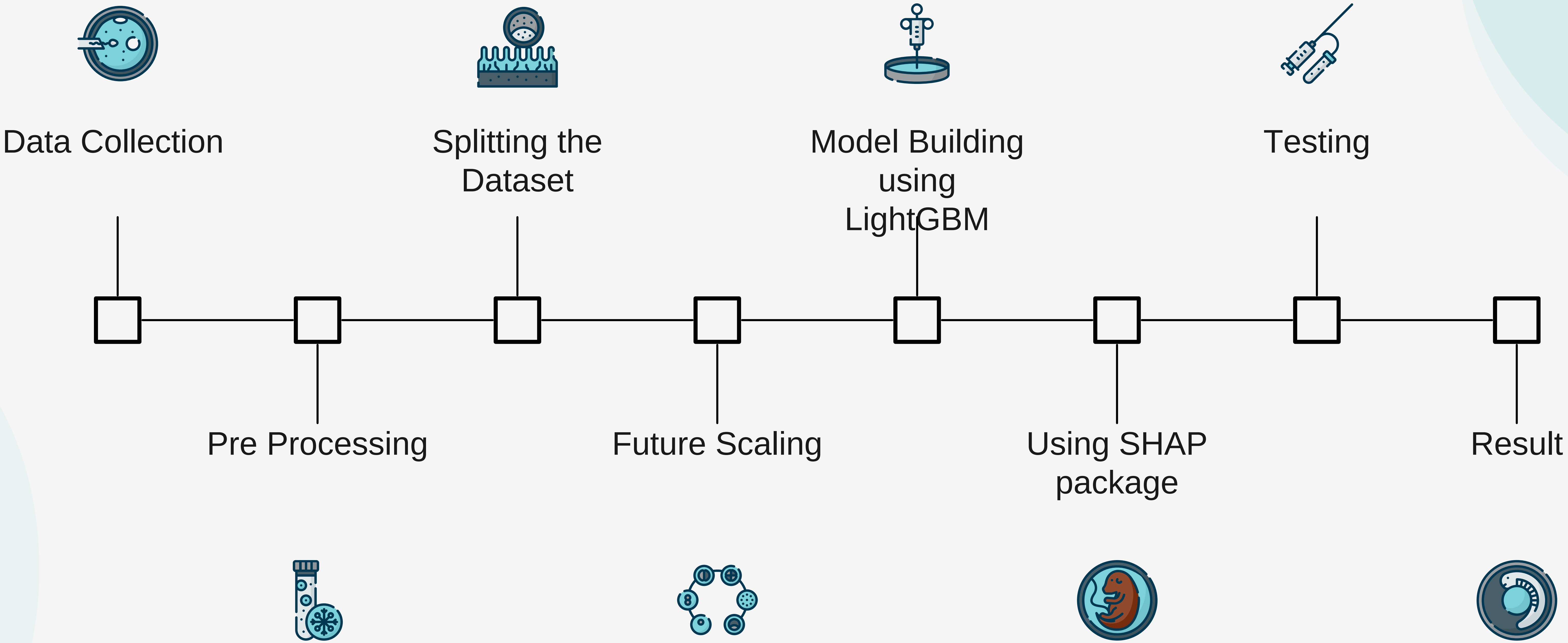
- Traditional methods of assessing the risks of air pollution during pregnancy rely on observational studies and statistical analysis, but lack real-time predictive capabilities. Integrating Artificial Intelligence, particularly using LightGBM and Shapley Additive Explanations, enhances this system by providing a deeper understanding of how different pollutants affect pregnant women and unborn children.

- This AI-driven model analyzes key features and offers explanations for predictions, empowering preventive strategies and public health interventions to address air pollution risks more comprehensively.

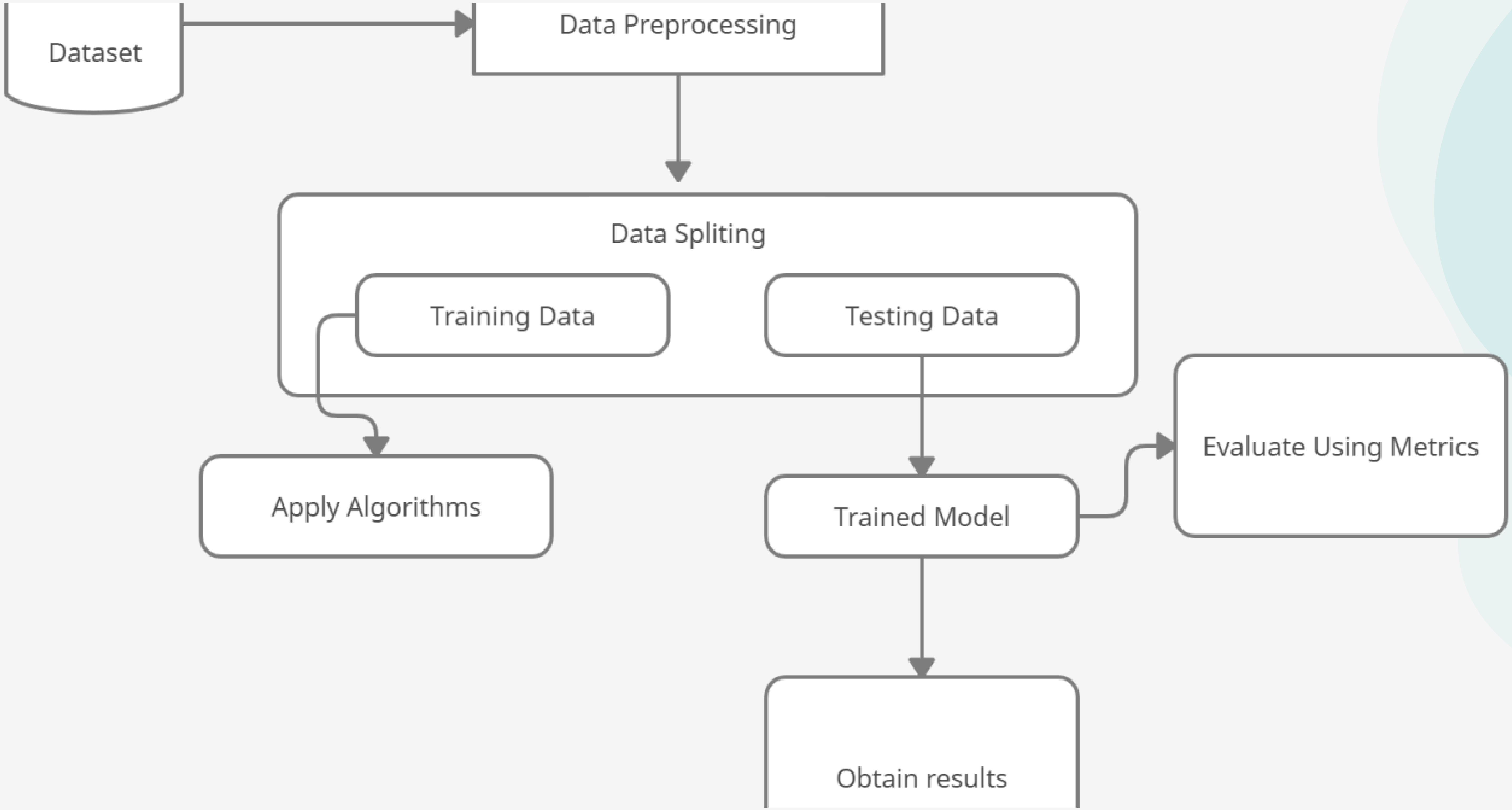
PROBLEM STATEMENT

- Air pollution poses significant risks during pregnancy, causing symptoms like coughing and respiratory issues while potentially harming the unborn child. This research focuses on understanding and predicting these effects, especially on vulnerable groups like the elderly, young children, and pregnant women. To address this challenge, we utilize LightGBM and Shapley Additive Explanations to analyze key features linked to air pollution and pregnancy, providing insights into potential risks and proposing strategies to mitigate harmful airborne emissions.

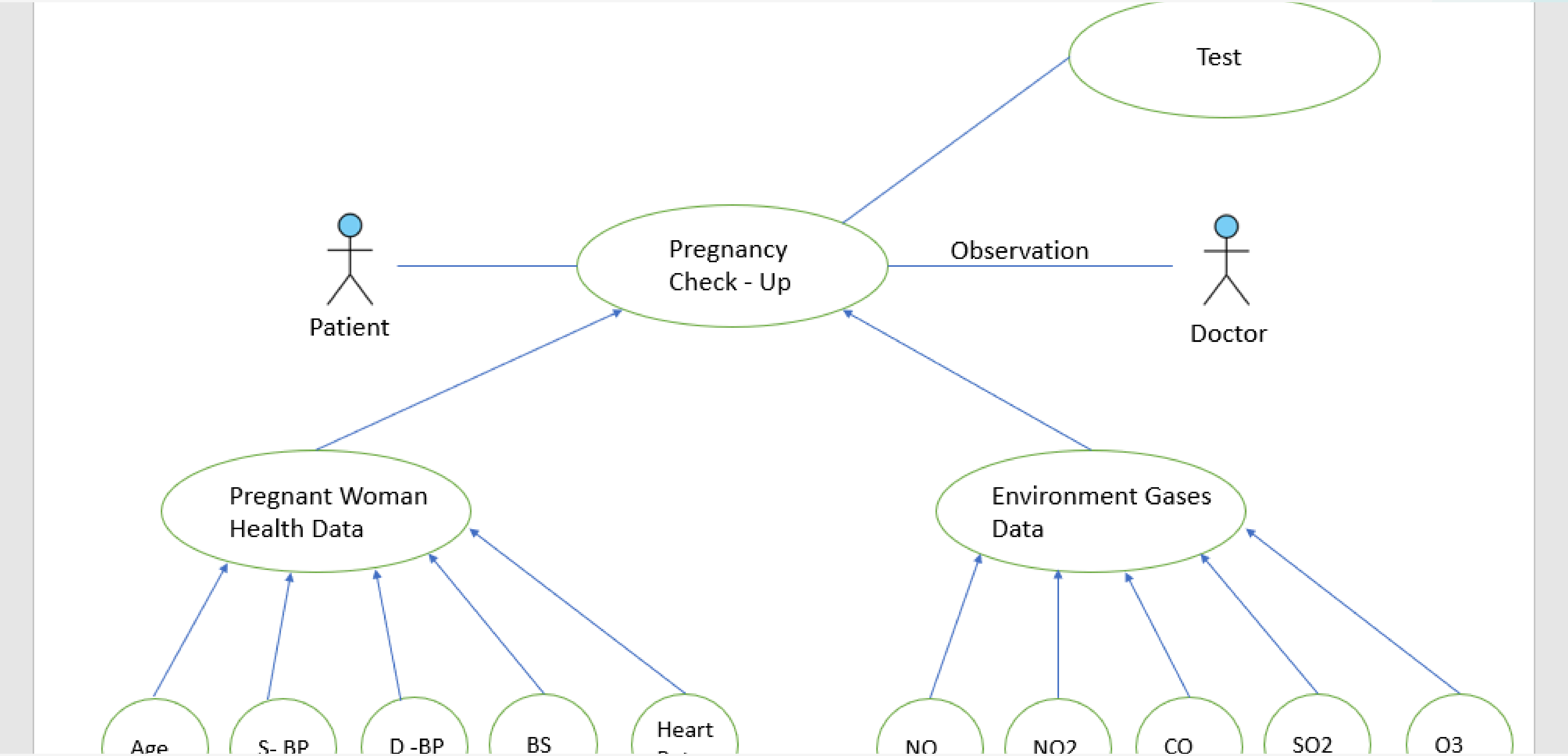
Methodology



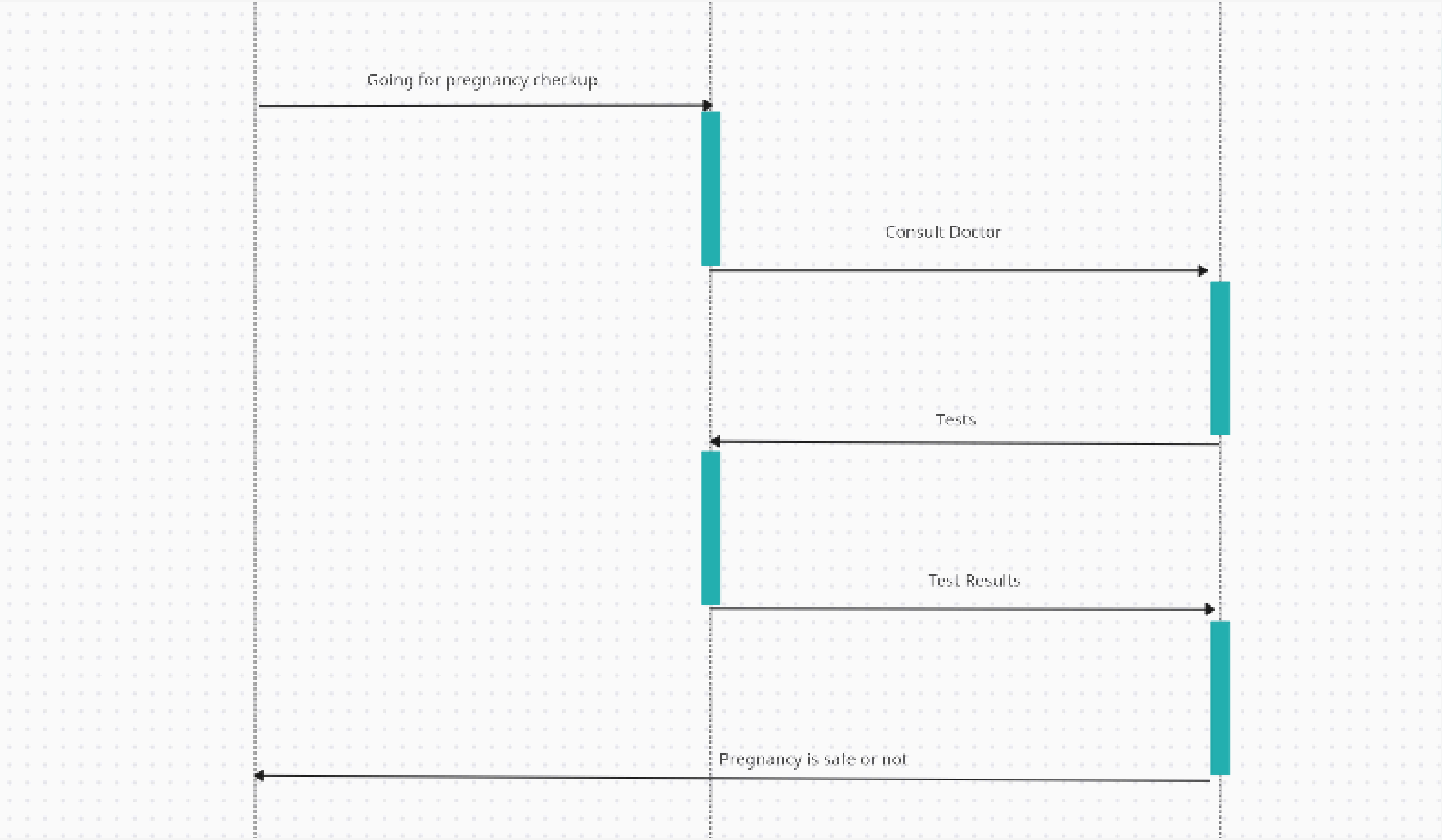
Architecture:



USE CASE DIAGRAM:



SEQUENCE DIAGRAM:



Implementation

Creating an **AI-based Potential Risks of Exposure to Air Pollution While Pregnant Model** involves several steps including Collecting the Pregnancy Women's Data, Creating a Dataset, Training a Model, Using the SHAP Method, and Creating a Flask Web Framework. Here's a breakdown of the explanation:

1. Collecting Data: In our project, we collected data on pregnant women from various healthcare facilities, research databases, and relevant literature to analyze the potential risks of air pollution exposure during pregnancy. Our dataset consists of 11 rows and 1014 columns, covering demographic details, medical histories, pollutant exposure levels, and pregnancy outcomes. This extensive dataset enables a thorough investigation using AI-based methodologies, aiming to improve understanding and inform prenatal care strategies.

Implementation

The dataset contains around 10 features and one target variable that predict whether the pregnant woman is at high risk or low risk. The features of the dataset are as follows:

Patient Data:

- **Age:** Age in years when a woman is pregnant.
- **Systolic BP:** Upper value of Blood Pressure in mmHg, another significant attribute during pregnancy.
- **Diastolic BP:-** The lower value of Blood Pressure in mmHg, another significant attribute during pregnancy.
- **Blood Glucose:-** Blood glucose levels are in terms of a molar concentration, mmol/L.
- **Heart Rate:-** A normal resting heart rate in beats per minute.

Implementation

Environment Gases Data:

- **NO:** Concentration of nitrogen monoxide in air measured in µg/m³
- **NO2:** Concentration of nitrogen dioxide in air measured in µg/m³
- **CO:** Concentration of carbon monoxide in air measured in µg/m³
- **SO2 :** Concentration of sulfur monoxide in air measured in µg/m³
- **O3 :** Concentration of ozone in air measured in µg/m³
- **Risk Level:** Predicted Risk Intensity Level during pregnancy considering the previous attributes.

Implementation

2. Creating a Dataset: Our dataset includes essential features such as Age, Blood Pressure, Blood Glucose, Heart Rate, and levels of air pollutants including NO, NO2, CO, SO2, and O3. Additionally, each entry is labeled with a Risk Level. This dataset enables precise analysis of the impact of air pollution on pregnant women's health.

3. Training a Model: After implementing preprocessing techniques, including null value removal and feature column transformation, we employed the **LightGBM Classifier** (LGBMClassifier) to train our model. Leveraging this advanced machine learning algorithm, we achieved an impressive accuracy score of 95.7%. This will accurately be predicting the impact of air pollution on pregnant women's health outcomes. The LGBMClassifier's robustness and efficiency make it ideal for handling complex datasets, extracting meaningful insights, and further validating our findings.

Implementation

4. SHAP Method: We utilized the SHAP (Shapley Additive exPlanations) method in our model building, accompanied by visualizations including the waterfall model, Force Plot, Stacked Force Plot, Heat Map, Absolute mean Shap, Beeswarm Plot, and Violin Plot. These visualizations provided precise insights into feature contributions and enhanced the interpretability of our model's predictions.

5. Creating a Flask Web Framework: We utilized Flask which is a lightweight, flexible, and efficient framework for our project. Its simplicity allowed for rapid development and deployment of a dynamic web application. With its extensive ecosystem of extensions, we easily implemented features like taking input like patient data, and environmental gasses data from the user and analyzing the result as to whether the pregnant woman is in danger or safe. also, we added the Contact Us page for the user queries.

Threshold Values:

Incorporating an image detailing threshold values of gases for air pollution exposure, enhancing visual clarity and understanding in our presentation.

µg/m3	Low levels	Moderate levels	High levels
SO2	20	20-100	100
CO	5	5-10	10
NH3	10	10-50	50
NO2	10	10-40	40
NOx	10	10-40	40

ppb/ppm	Low levels	Moderate levels	High levels
SO2	50	50-100	100
CO (ppm)	1	1-5	5
NH3	20	20-100	100
NO2	20	20-100	100
NOx	20	20-100	100

Sample Code:

PROJECT-Copy1.ipynb

C: > Users > RAMISETTI NARASIMHA > Desktop > final year project review 1 > pregnency_model_implementation > PROJECT-Copy1.ipynb > import numpy as np

+ Code + Markdown | Run All Restart Clear All Outputs | Variables Outline ... tf (Python 3.10.9)

```python
import numpy as np
import pandas as pd
from sklearn.model_selection import train_test_split
from lightgbm import LGBMClassifier
import shap
from sklearn.metrics import roc_auc_score
from sklearn.metrics import roc_curve
from math import sqrt
from sklearn.metrics import confusion_matrix
from sklearn.metrics import classification_report
import matplotlib.pyplot as plt
```

[7] Python

```python
df=pd.read_csv(r"C:\Users\RAMISETTI NARASIMHA\Desktop\final year project review 1\pregnency_model_implementation\project_data.csv")
df
```

[8] Python

	Age	SystolicBP	DiastolicBP	Blood Glucose	HeartRate	NO	NO2	CO	SO2	O3	Risk Level
0	25.0	130.0	80.0	15.0	86.0	100.27	61.29	2.48	5.36	58.24	High Risk
1	35.0	140.0	90.0	13.0	70.0	87.94	51.83	2.02	25.60	62.02	High Risk
2	29.0	90.0	70.0	8.0	80.0	74.45	52.92	1.98	14.75	65.55	High Risk
3	30.0	140.0	85.0	7.0	70.0	25.13	48.92	1.89	11.84	73.28	High Risk
4	35.0	120.0	60.0	6.1	76.0	14.33	43.47	1.61	10.89	67.89	High Risk
...	...	...	...	...	...	...	...	...	...	...	...
1009	22.0	120.0	60.0	15.0	80.0	21.42	29.38	1.21	10.04	43.73	Low Risk
1010	55.0	120.0	90.0	18.0	60.0	3.49	28.91	0.96	12.39	53.56	High Risk
1011	35.0	85.0	60.0	19.0	86.0	24.93	69.34	1.58	13.23	36.64	High Risk
1012	43.0	120.0	90.0	18.0	70.0	10.79	56.23	1.39	18.31	53.06	High Risk
1013	32.0	120.0	65.0	6.0	76.0	5.75	35.27	1.03	15.56	37.27	Low Risk

1014 rows × 11 columns

Sample Code:

PROJECT-Copy1.ipynb

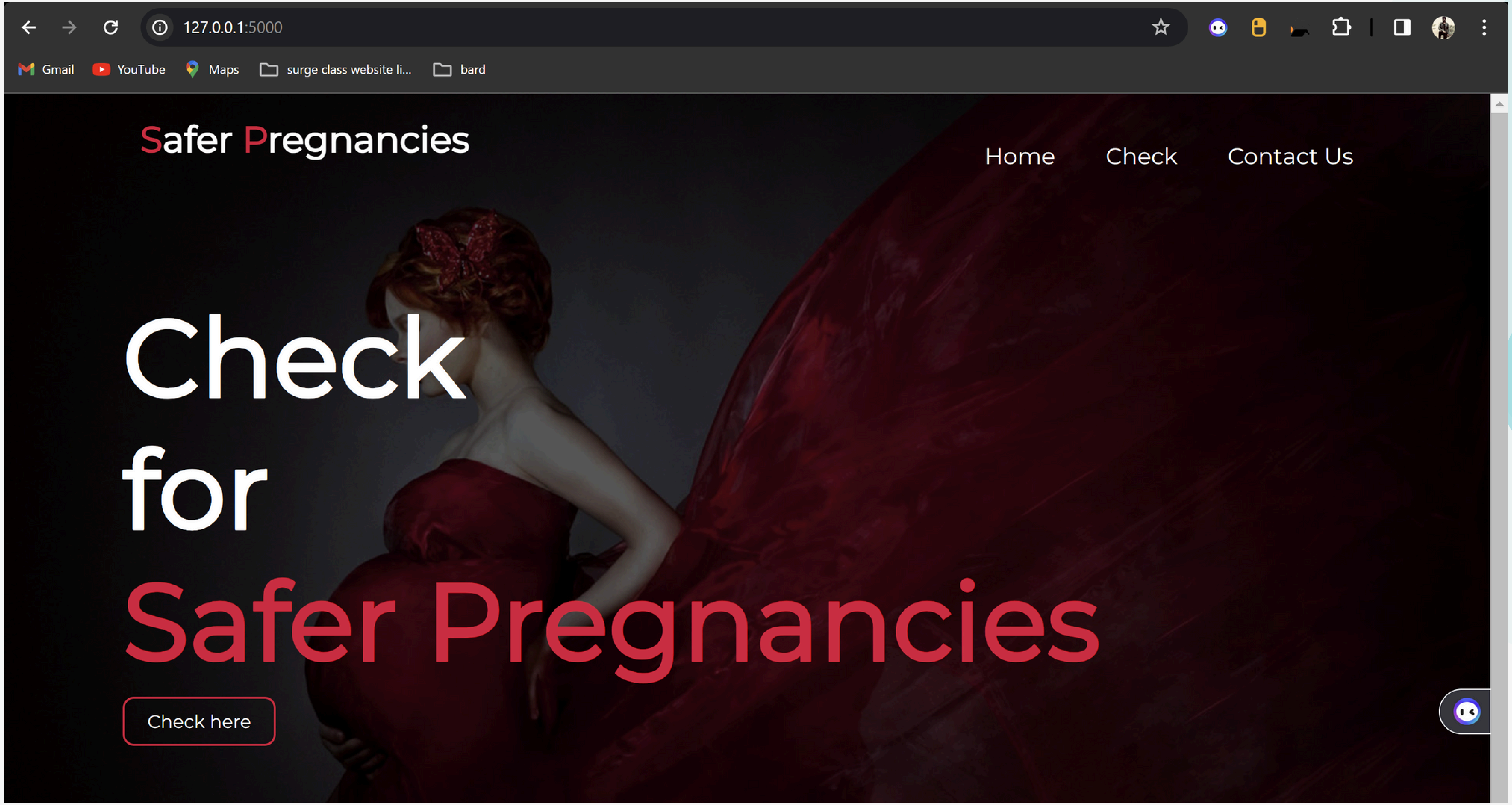
C: > Users > RAMISETTI NARASIMHA > Desktop > final year project review 1 > pregnency_model_implementation > PROJECT-Copy1.ipynb > import numpy as np

+ Code + Markdown | Run All Restart Clear All Outputs | Variables Outline ··· tf (Python 3.10.9)

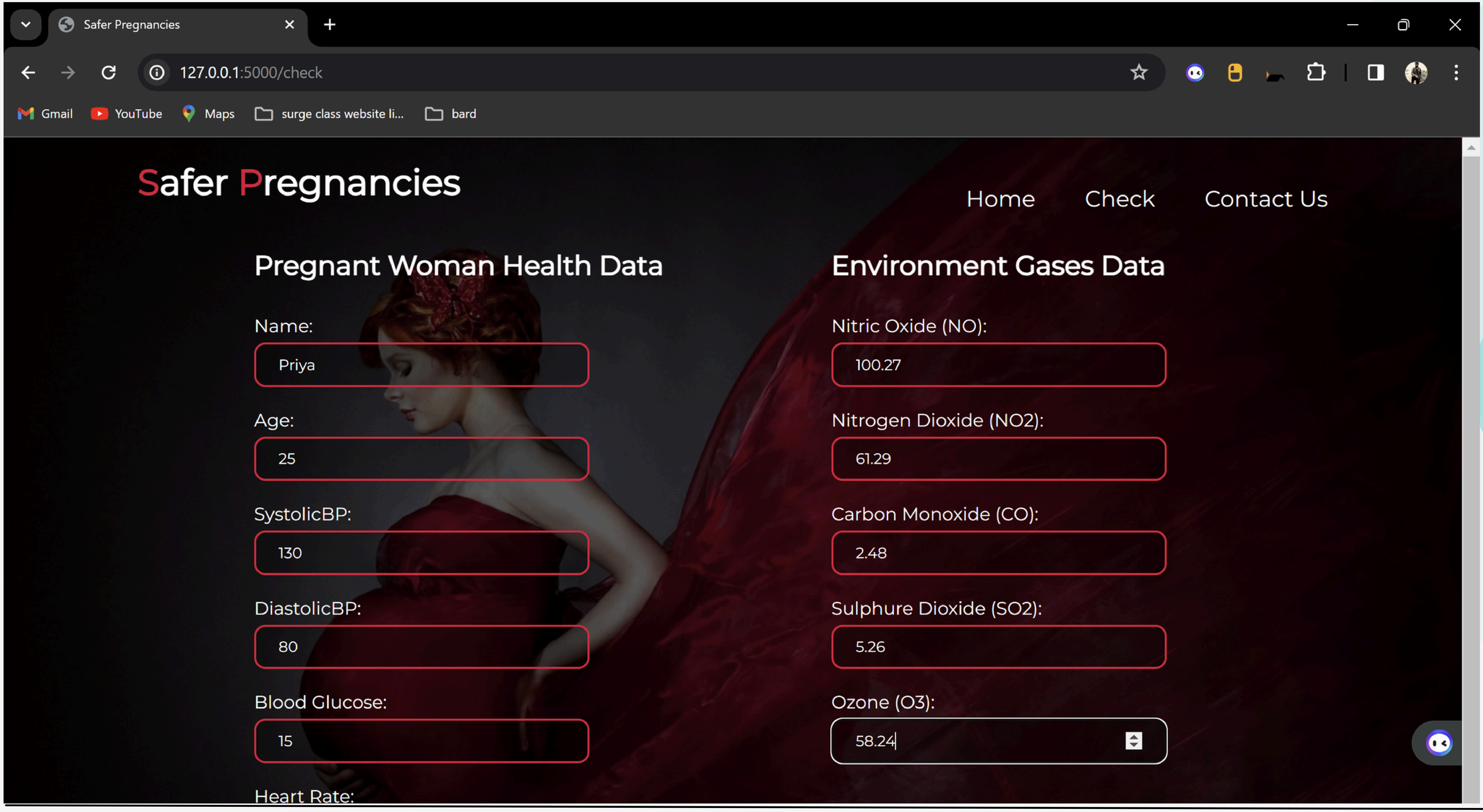
Evaluation Metrics

```python
actual = y_test
predicted = y_pred
matrix = confusion_matrix(actual,predicted, labels=[1,0],sample_weight=None, normalize=None)
print('Confusion matrix : \n', matrix)
tp, fn, fp, tn = confusion_matrix(actual,predicted,labels=[1,0]).reshape(-1)
print('Outcome values : \n', tp, fn, fp, tn)
C_Report = classification_report(actual,predicted,labels=[1,0])
print('Classification report : \n', C_Report)
sensitivity = round(tp/(tp+fn), 3);
specificity = round(tn/(tn+fp), 3);
accuracy = round((tp+tn)/(tp+fp+tn+fn), 3);
balanced_accuracy = round((sensitivity+specificity)/2, 3);
precision = round(tp/(tp+fp), 3);
f1Score = round((2*tp/(2*tp + fp + fn)), 3);
mx = (tp+fp) * (tp+fn) * (tn+fp) * (tn+fn)
MCC = round(((tp * tn) - (fp * fn)) / sqrt(mx), 3)
print('Accuracy :', round(accuracy*100, 2),'%')
print('Precision :', round(precision*100, 2),'%')
print('Recall :', round(sensitivity*100,2), '%')
print('F1 Score :', f1Score)
print('Specificity or True Negative Rate :', round(specificity*100,2), '%'  )
print('Balanced Accuracy :', round(balanced_accuracy*100, 2),'%')
print('MCC :', MCC)

print('roc_auc_score:', round(roc_auc_score(y_test, y_pred), 3))

logit_roc_auc = roc_auc_score(y_test, y_pred)
fpr, tpr, thresholds = roc_curve(y_test,model.predict_proba(x_test)[:,1])
plt.figure()
plt.plot(fpr, tpr, label= 'Classification Model' % logit_roc_auc)
plt.plot([0, 1], [0, 1],'r--')
plt.xlim([0.0, 1.0])
plt.ylim([0.0, 1.05])
plt.xlabel('False Positive Rate')
plt.ylabel('True Positive Rate')
plt.title('Receiver operating characteristic')
```

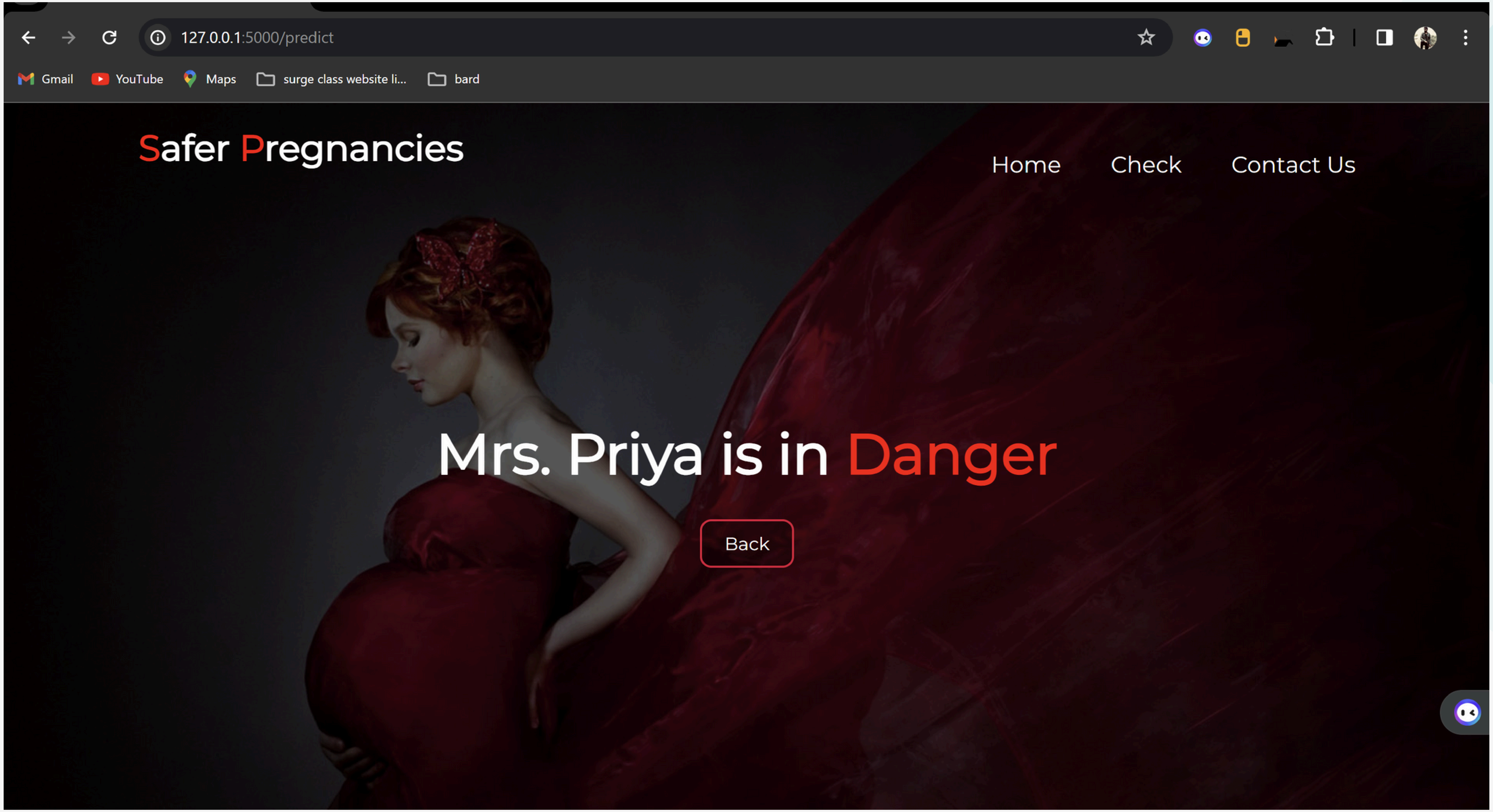
Sample Input:



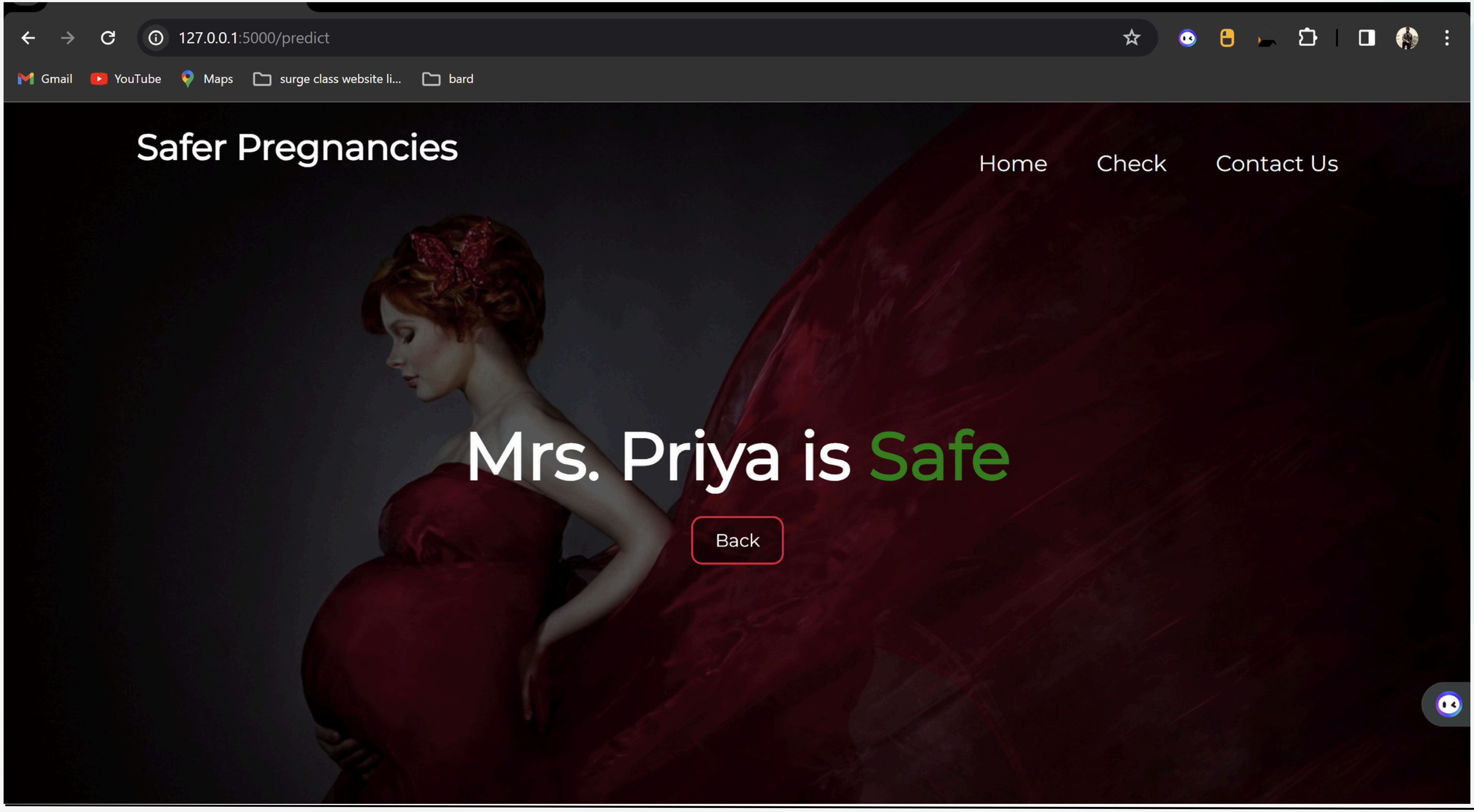
Sample Input:



Sample Input:



Sample Input:



Technical Stack Used

1. SHAP Method: We utilized the SHAP (Shapley Additive exPlanations) method in our model building, accompanied by visualizations including the waterfall model, Force Plot, Stacked Force Plot, Heat Map, Absolute mean Shap, Beeswarm Plot, and Violin Plot. These visualizations provided precise insights into feature contributions and enhanced the interpretability of our model's predictions.

2. NumPy (Numerical Python): NumPy is a fundamental package for scientific computing with Python, enabling efficient manipulation and computation of multi - dimensional arrays. It is utilized in the project for various mathematical operations and data manipulation tasks, such as converting images to arrays and performing array-based operations during image processing.

Technical Stack Used

3. Pickle: Pickle is a Python module used for serializing and deserializing Python objects. In this project, pickle may be used to save trained machine learning models to disk, allowing for easy storage and retrieval of the model for future use without the need to retrain it every time

4. NumPy (Numerical Python): NumPy is a fundamental package for scientific computing with Python, enabling efficient manipulation and computation of multi-dimensional arrays. It is utilized in the project for various mathematical operations and data manipulation tasks, such as converting images to arrays and performing array-based operations during image processing.

Results:

EVALUATION METRICS:

```
+ Code  + Markdown | Run All  Clear All Outputs | Outline ...                Select Kernel
  39  plt.show()
[20]                                                                          Python

... Confusion matrix :
     [[120   8]
     [  3 123]]
    Outcome values :
     120 8 3 123
    Classification report :
                   precision    recall  f1-score   support

               1       0.98      0.94      0.96       128
               0       0.94      0.98      0.96       126

        accuracy                           0.96       254
       macro avg       0.96      0.96      0.96       254
    weighted avg       0.96      0.96      0.96       254

    Accuracy : 95.7 %
    Precision : 97.6 %
    Recall : 93.8 %
    F1 Score : 0.956
    Specificity or True Negative Rate : 97.6 %
    Balanced Accuracy : 95.7 %
```

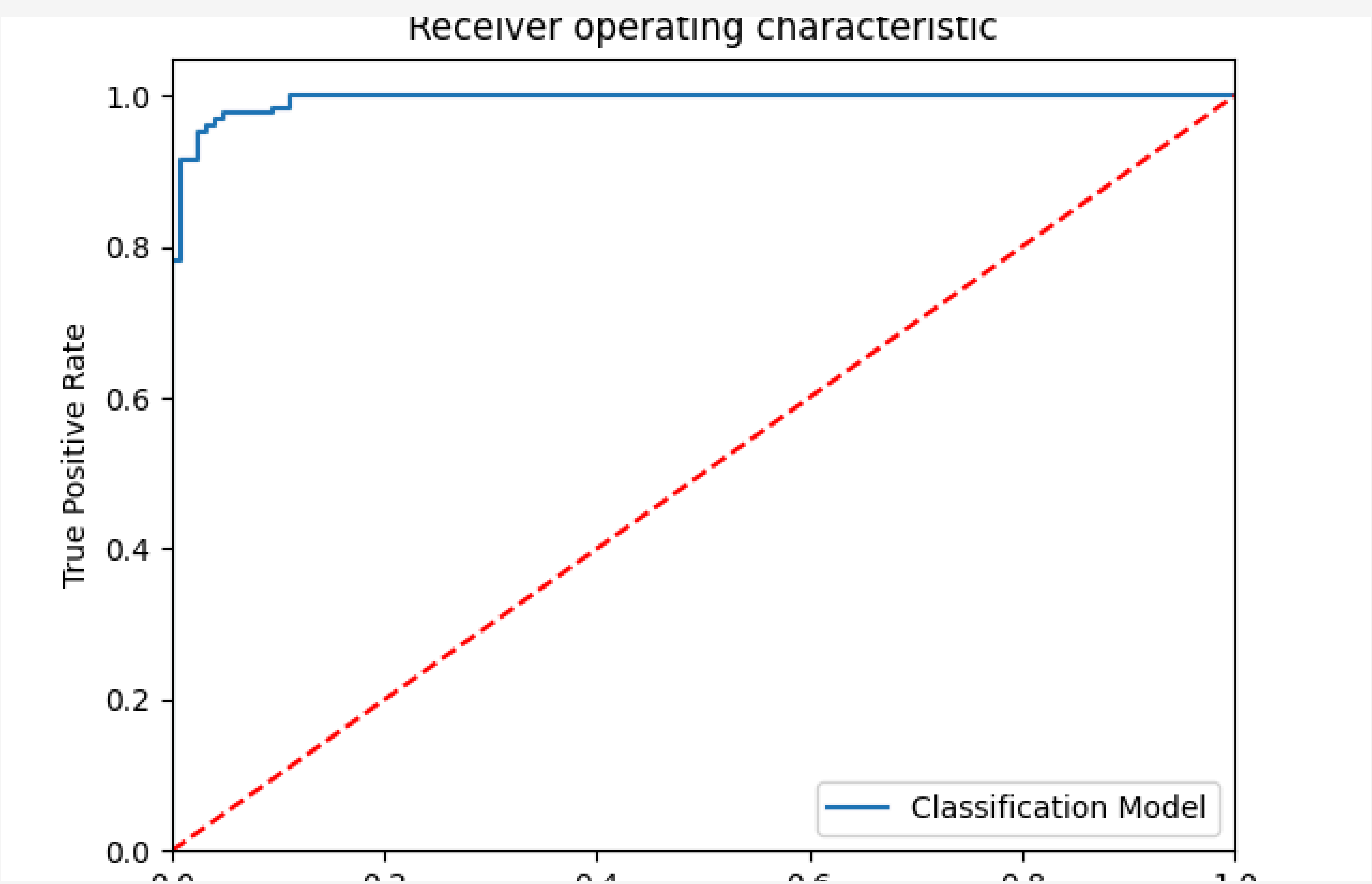
Results:

- Our model using LightGBM yielded an accuracy score of 95.7%, highlighting its effectiveness in predicting the health impact of air pollution on pregnant women.

Model	Accuracy	Precision	Recall	Specificity	F1 Score
Logistic Regression	83.5	84.7	82.0	84.9	0.833
Decision Tree	92.5	93.6	91.4	93.7	0.925
Random Forest	92.1	95.8	88.3	96.0	0.919
Light GBM	**95.7**	**97.6**	**93.8**	**97.6**	**0.956**
Extra Trees	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

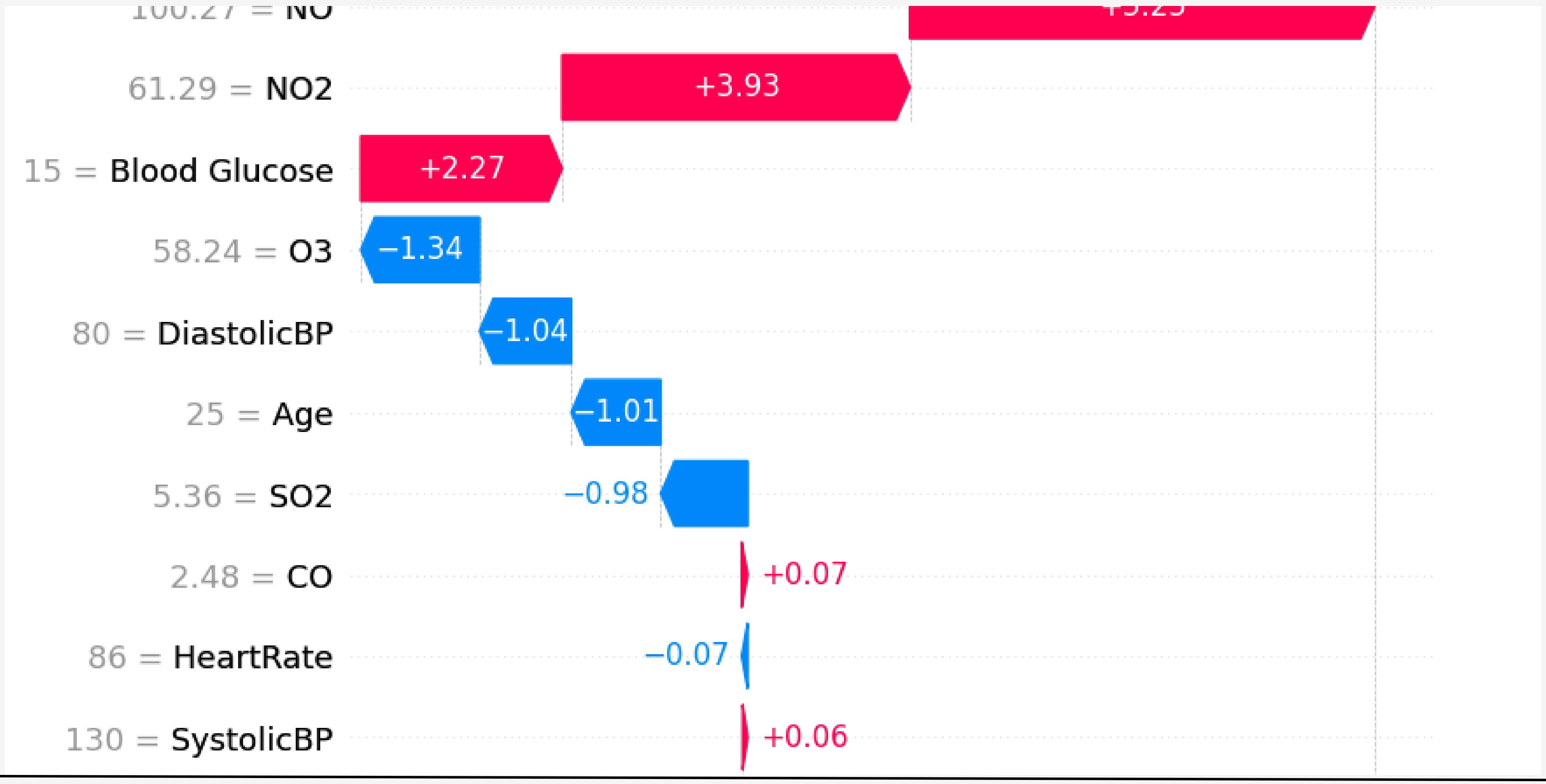
Results:

AUC – ROC CURVE



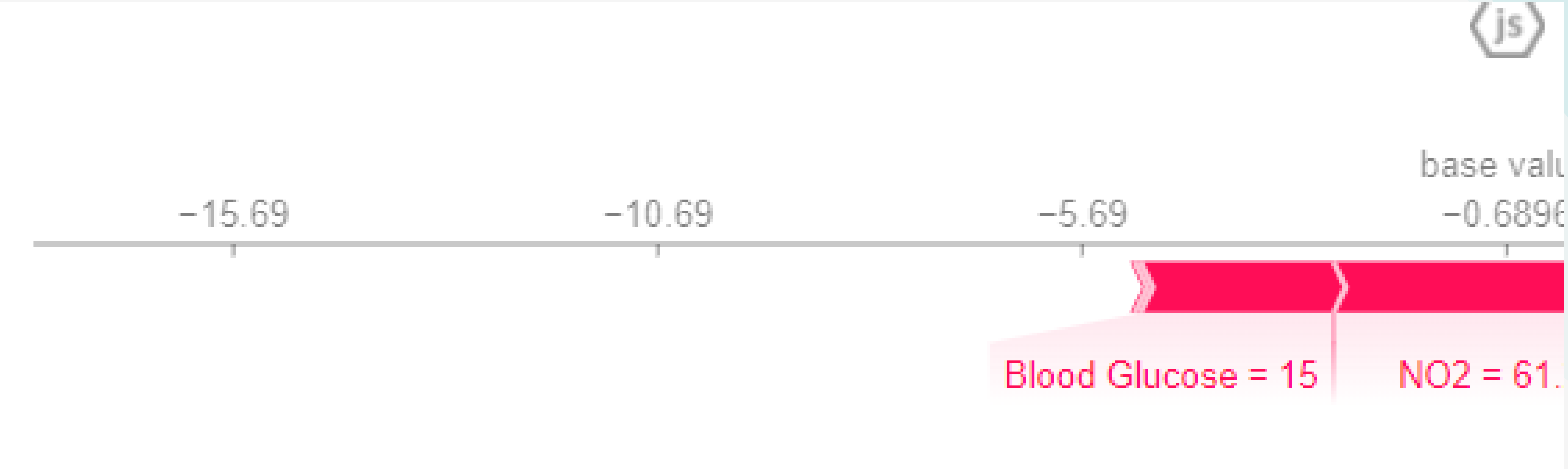
Results: SHAP PLOTS

WATERFALL PLOT: The waterfall plot is designed to visually display how the SHAP values (evidence) of each feature move the model output from our prior expectation under the background data distribution to the final model prediction given the evidence of all the features.



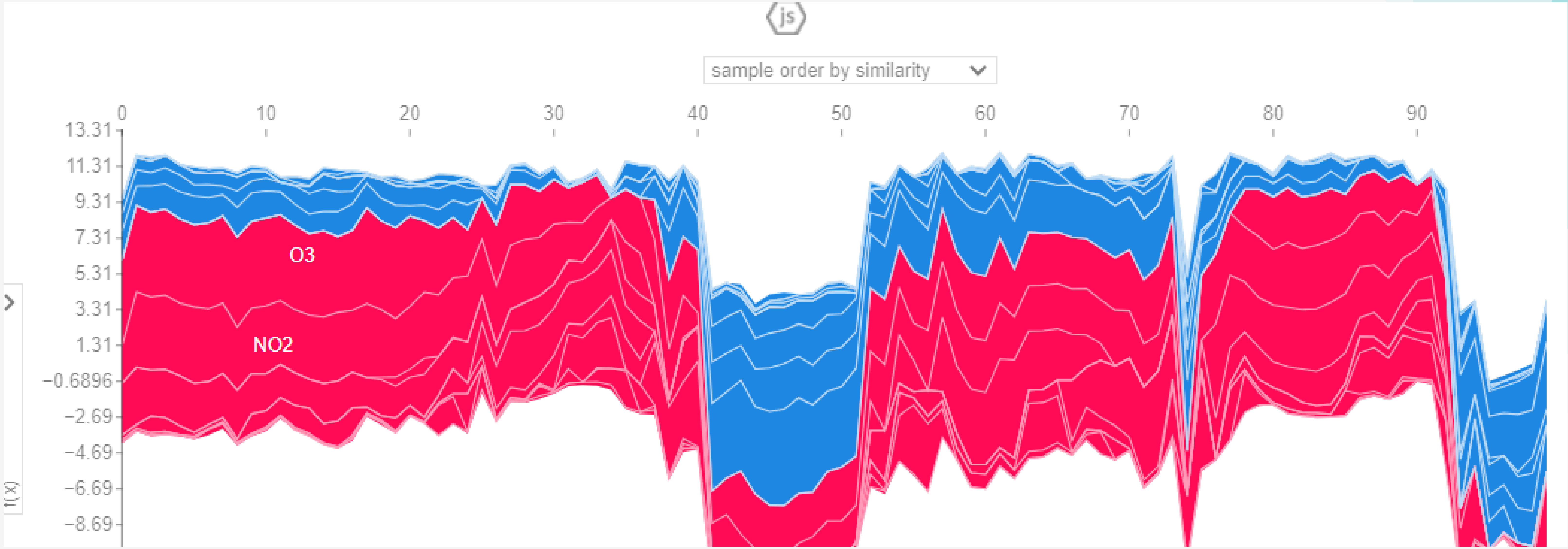
Results: SHAP PLOTS

FORCE PLOT: SHAP force plots show the features from left to right, with the positive contributions on the right and the negative contributions on the left. Each feature's contribution is represented by a bar, and the length and direction of the bar indicate the magnitude and direction of the feature's effect on the prediction, respectively.



Results: SHAP PLOTS

Stacked FORCE PLOT: The stacked force plot is particularly useful for examining misclassified instances and gaining insights into the factors driving those misclassifications. This allows for a deeper understanding of the model's decision-making process and helps pinpoint areas that require further investigation or improvement.



Conclusion:

In conclusion, our project has successfully demonstrated the efficacy of utilizing advanced methodologies such as the LightGBM Classifier and SHAP (Shapley Additive exPlanations) method to analyze the potential risks associated with air pollution exposure during pregnancy. Leveraging a comprehensive dataset and sophisticated modeling techniques, we achieved an impressive accuracy score of 95.7%, highlighting the reliability and robustness of our approach. Through the integration of the Flask web framework, we developed a user-friendly platform capable of providing real-time analysis and insights. This project not only enhances our understanding of the intricate relationship between air pollution and maternal health but also offers valuable insights for informing prenatal care strategies and mitigating risks for pregnant women. Moving forward, our findings hold significant implications for public health policies aimed at safeguarding maternal and fetal well-being in areas affected by air pollution.

Future Work:

For future work, we aim to enhance our project by incorporating additional features that provide valuable insights for pregnant women's health. After executing the model, we plan to display detailed information on which gases are affecting pregnant women more prominently on the result webpage. This feature will offer users a clear understanding of the specific pollutants posing the greatest risks, enabling informed decision-making and targeted interventions. Additionally, we intend to include comprehensive data on informing prenatal care strategies and mitigating risks for pregnant women directly on the result webpage. By integrating these enhancements, we aspire to empower users with actionable information, further advancing the effectiveness of our platform in promoting maternal and fetal well-being amidst air pollution challenges.

References:

- [1] Alexandra Grippo, Ajay A.”Air Pollution exposure during pregnancy and spontaneous abortions and stillbirth” published on 25-09-2022.
- [2] Prafulla Shriyan, Deepa Ravi, Giridhara R Babu, Yamuna A “Ambient and Indoor Air Pollution in Pregnancy and the risk of Low birth weight and Ensuing Effects in Infants: A Cohort Study in Bangalore, South India” published on 19-10-2021.
- [3] “Environmental Exposure during pregnancy: influence on prenatal development and early life: a comprehensive review” published in 2021.

Thanks