MATERNAL MORTALITY

1st Neha Rana
Master in Data Science
nrana4@ur.rochester.edu
University Of Rochester

2nd Ashika Kotia
Master in Data Science
akotia@ur.rochester.edu
University Of Rochester

***Abstract*—This Project is a comprehensive study aimed at understanding the causes and contributing factors to maternal mortality. The primary objective of this project is to identify key areas where interventions can significantly reduce the rate of maternal deaths.**

I. Introduction

Maternal mortality is the death of a woman during pregnancy or within 42 days of the termination of pregnancy. Despite significant progress in healthcare and medical advancements, maternal mortality remains a critical concern, particularly in low- and middle-income countries. Maternal death rates are rising again in many countries, yet this alarming fact has gone largely unacknowledged in the medical and lay press. In 2020, one woman died every 2 min from preventable causes related to pregnancy. This statistic represents approximately 800 women dying daily—a maternal mortality ratio (MMR) of 223 maternal deaths per 100,000 live births, which is far removed from the UN Sustainable Development Goal target of 3.1 to reduce the global MMR to less than 70 deaths per 100,000 live births by 2030. Between 2000 and 2020 the global MMR declined by 34·8 percent (from 342 deaths per 100,000 live births to 223 deaths per 100,000 live births). During the period 2016–20, MMRs stagnated in 133 countries, and increased substantially in 17, mainly in Western Europe, North America, Latin America, and the Caribbean. Between 2000 and 2020, eight countries recorded substantial increases in MMRs: Venezuela (182.8), Cyprus (107), Greece (101.1), the USA (77.9), Mauritius (62.1), Puerto Rico (55·9), Belize (51·3), and the Dominican Republic (36·0; figure A). In 2021, differences between countries mask the true burden of mortality in countries with the highest MMRs. In three sub-Saharan African countries, the MMR in 2020 exceeded 1000 deaths per 100000 live births: South Sudan (1223), Chad (1063), and Nigeria (1047; figure B). Nigeria alone recorded 82000 deaths in 2020, representing 28.5 of global maternal deaths, while India, the Democratic Republic of the Congo, and Ethiopia each recorded over 10000 deaths The U.S. is considered an outlier among wealthy nations in maternal health care where the MMR increased from 20.1 deaths per 100,000 live births in 2019 to 23.8 deaths per 100,000 live births in 2020, and 32.9 deaths per 100,000 live births with significantly higher mortality rates recorded among Black women than among White and Hispanic women. MMRs increased across all age groups in the USA between 2020 and 2021, with a strong age gradient ranging from 20.4 deaths per 100000 live births in women younger than 25 years to 138 deaths per 100000 live births in those older than 40 years.

Identify applicable funding agency here. If none, delete this.

II. Related Work

In the field of maternal mortality research, several prominent organizations contribute significantly to the global understanding and improvement of maternal health outcomes. The World Health Organization (WHO) is a key player, offering comprehensive reports and data analysis on maternal mortality across various regions and countries, highlighting trends and disparities [4]. Similarly, UNICEF plays a crucial role by conducting extensive research focused on maternal and child health, utilizing advanced statistical methods to derive actionable insights from their data [3]. The World Bank, through its Open Data initiative, provides valuable datasets that explore the intricate relationship between health metrics, including maternal mortality, and economic factors. In the United States, the Centers for Disease Control and Prevention (CDC) are instrumental in advancing our understanding of maternal health. They employ epidemiological and statistical techniques to analyze health outcomes both domestically and internationally [2]. The Bill and Melinda Gates Foundation, with its commitment to improving global maternal health, channels substantial resources into research, interventions, and partnerships. Their Maternal, Newborn, and Child Health Program advocates for a comprehensive approach, addressing the interconnected health needs of mothers and children. This holistic strategy aims to foster a healthier future globally through strategic investments and collaborations [1]. Lastly, the National Institutes of Health (NIH) in the United States significantly contributes to this field through its federal research initiatives, particularly in areas concerning age and race-related health disparities [7].

III. Data Source

In our study of maternal health, we have utilized data from several pivotal sources, including the World Health Organization (WHO) [4], UNICEF, the World Bank, and two key databases from the Centers for Disease Control and

Prevention (CDC) - CDC's WONDER and the Pregnancy Mortality Surveillance System. The WHO offers a global perspective on maternal mortality, providing detailed data on trends and regional differences. UNICEF's contributions are centered around maternal and child health, employing sophisticated statistical methodologies [3]. The World Bank's Open Data initiative enriches our research with its insights into the economic aspects influencing maternal health. From the CDC, the WONDER database provides extensive epidemiological data on U.S. maternal health [9], while the Pregnancy Mortality Surveillance System specifically tracks and analyzes pregnancy-related deaths, offering vital information on the causes and patterns of maternal mortality in the United States [9]. These combined data sources form a comprehensive base for a thorough and nuanced understanding of maternal health issues globally and domestically.

IV. Data Preprocessing

In our project, a significant emphasis was placed on the integration and harmonization of data from our top sources, notably the CDC's WONDER database and the CDC's Pregnancy Mortality Surveillance System. This process involved a meticulous mapping of state-wise maternal death counts obtained from the CDC WONDER database to the more detailed individual-level data from the CDC's Pregnancy Mortality Surveillance System. After that, we tackled missing values, a common challenge in healthcare datasets. Depending on their prevalence and significance, we employed strategies like mean or median imputation for continuous variables and mode imputation or the creation of a separate category for missing values in categorical variables. Next, we encoded categorical variables through one-hot encoding or label encoding, translating qualitative data into a numerical format that the GBM model could interpret. We also standardized or normalized continuous variables to ensure uniformity in scale, which is particularly important for models sensitive to variable scaling. Furthermore, we identified and addressed outliers, either by removing them or using transformation techniques to minimize their impact.

V. Methodology

In our project, we started with a multifaceted exploration of the maternal mortality rate (MMR) across the globe, with a specific emphasis on the United States, where MMR is paradoxically rising in contrast to other developed nations. We began with the meticulous gathering of data from an array of reliable and authoritative sources. This foundational step was crucial in setting the stage for our comprehensive analysis. We utilized Tableau, a robust tool for data visualization, to conduct a descriptive analysis of this data. This phase was instrumental in unraveling the complex layers of MMR, highlighting the primary causes of maternal deaths in the USA, and revealing regional trends and disparities. As our investigation deepened, we integrated a pivotal dataset from the Centers for Disease Control and Prevention (CDC) – the Pregnancy-Related Mortality Surveillance System. This dataset proved invaluable, offering a wealth of information on various factors that could influence maternal mortality. Key variables in this dataset included weighted percentages of multivitamin use, pre-pregnancy weight status (categorized as overweight, obese, or underweight), substance use, health insurance coverage, and the initiation of prenatal care in the first trimester. These variables opened new avenues of exploration and allowed us to delve deeper into the potential determinants of maternal mortality. Recognizing the complexity of the data and the limitations of traditional analytical methods like linear and logistic regressions, particularly their inadequacy in handling the non-linearity and high dimensionality typical of health-related datasets, we turned to a more sophisticated modeling approach. The Gradient Boosting Machine (GBM) model emerged as our method of choice. This model is particularly adept at analyzing complex datasets, capable of uncovering both linear and non-linear relationships. Our application of the GBM model yielded a mean squared error (MSE) of 177.49, a testament to its predictive accuracy. It also helped us identify and rank various factors in terms of their impact on maternal mortality, such as flu shot status, insurance coverage, and substance use, including specific variables like 'Hookah use in the last 2 years' and 'No insurance - Upper 95 Confidence Interval'. These insights were crucial in understanding the multifarious nature of factors influencing maternal mortality. In addition to the GBM model, we employed linear regression to further dissect the impacts of lifestyle and mental health factors, such as smoking, drinking, depression, and intimate partner violence (IPV), on maternal deaths. This analysis revealed that these factors accounted for approximately 34 of the variation in the maternal death rate per 100,000 individuals. This finding was significant in highlighting the role of personal and mental health factors in maternal mortality.

A. Gradient Boosting Machine (GBM) model

In our project, we employed the Gradient Boosting Machine (GBM) model, a sophisticated and powerful machine learning algorithm that belongs to the family of ensemble learning methods. GBM operates by constructing a series of decision trees in a sequential manner, where each tree corrects the errors made by its predecessors. This iterative process, known as boosting, effectively combines the predictions from multiple models to create a more accurate and robust predictive tool. The use of gradient descent in GBM, a method for minimizing the loss (error), further refines its predictions. The choice of GBM for our dataset was driven by its ability to handle complex, non-linear relationships inherent in healthcare data, making it exceptionally suitable for analyzing the intricate patterns and factors influencing maternal mortality. Additionally, the GBM model is well-equipped to manage the high-dimensional space of our dataset, efficiently selecting significant features and interactions, thereby enhancing the model's performance and interpretability. Its robustness to different types of data, ability to handle missing values, and resistance to overfitting, especially in complex datasets like ours, make it an ideal choice for providing deep insights into maternal health outcomes.

B. Linear regression

Linear regression is a statistical method used to model the relationship between a dependent variable and one or more independent variables, assuming a linear connection between them. This approach is valued for its simplicity and interpretability, making it effective for understanding the strength and direction of relationships between variables. In our project, following the advanced analysis with the Gradient Boosting Machine (GBM) model, we utilized linear regression for a focused examination of specific features identified by the GBM's feature selection. Linear regression provided a straightforward framework to assess the direct impact of these key variables on maternal mortality. This method allowed for clear hypothesis testing and quantifiable insights into how these selected factors influenced maternal mortality rates. By complementing the GBM's complex analysis with linear regression, we achieved a comprehensive and nuanced understanding of the determinants of maternal mortality, enhancing the overall validity of our findings.

VI. Experiment

We performed analysis on a dataset to understand the trends in health behaviors and statuses among pregnant women from 2016 to 2021. Our analysis began by plotting the yearly trends of various health indicators, including underweight, overweight, obesity, multivitamin use, hookah use, and e-cigarette use before and during pregnancy. These indicators were plotted as weighted percentages to account for the population distribution within the dataset. We observed a decrease in the average underweight percentage over the years, suggesting a potential improvement in prenatal health or changes in population health status. Conversely, the average overweight percentage and obese percentage showed an upward trend, indicating a growing concern for maternal health that may impact MMR. The use of e-cigarettes before pregnancy was seen to increase steadily, with a more pronounced rise during pregnancy, highlighting a worrying trend in maternal health behaviors. Multivitamin use showed a significant positive trend, which could be associated with increased health awareness or improved health policies. Hookah use, on the other hand, displayed a general decline, suggesting a possible shift in societal habits or effective public health interventions. For each health behavior or status indicator, we computed the weighted percentages using the dataset's stratification factors to ensure our findings were reflective of the underlying population structure. The analysis was conducted using statistical software that enabled the handling of complex survey data, accounting for stratification, clustering, and weighting to produce accurate estimates representative of the population. These trends provided us with a comprehensive view of the changing patterns in health behaviors and statuses that could influence maternal health outcomes and informed further in-depth analysis using advanced statistical models.

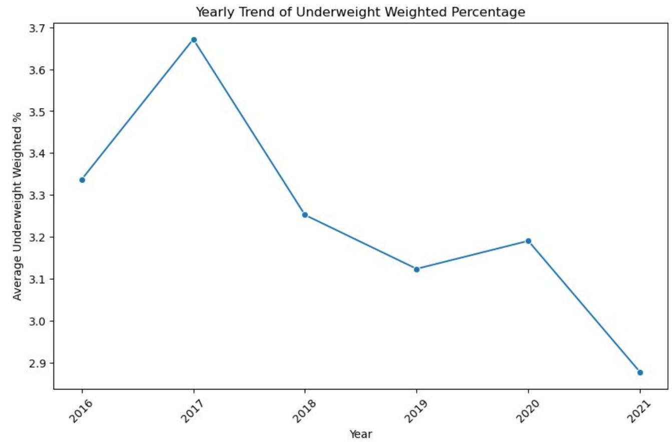

Fig. 1. Trend of Underweight

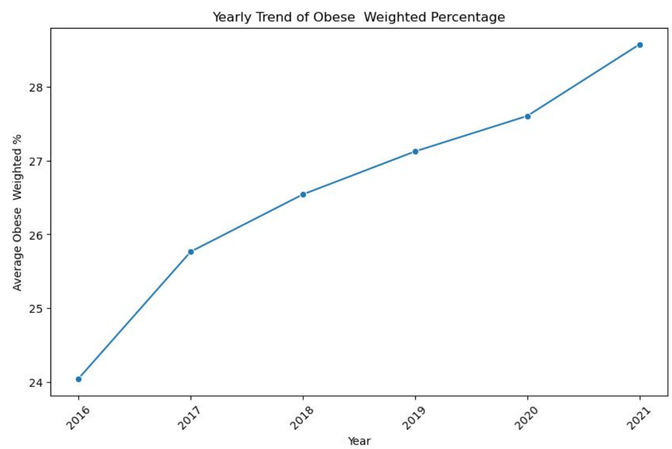

Fig. 2. Trend of Overweight

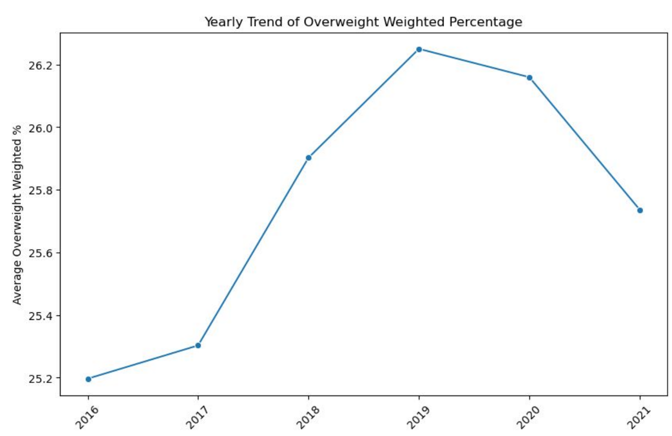

Fig. 3. Trend of Obese

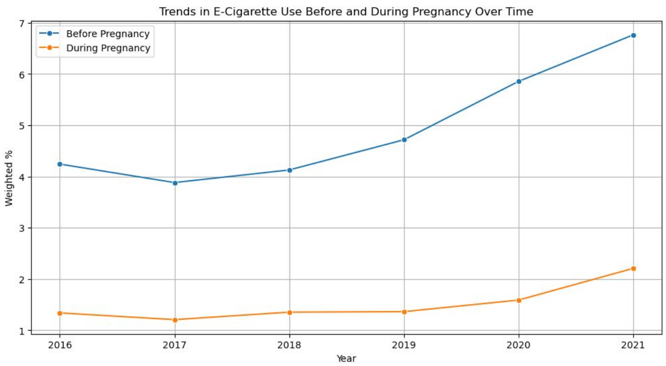

Fig. 4. Trend in E- Cigarette Use

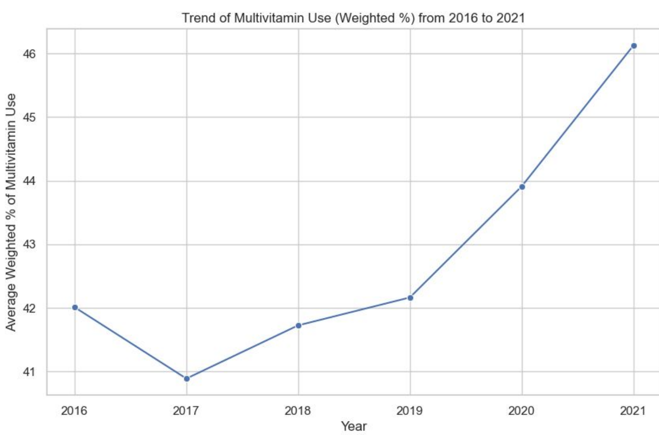

Fig. 5. Trend of Multivitamin Use

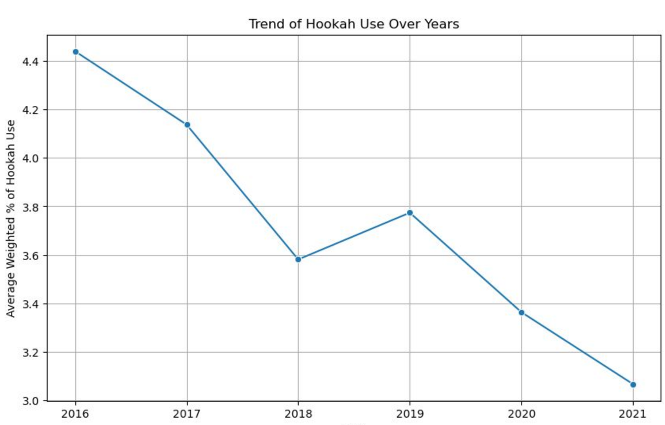

Fig. 6. Trend of Hookah Use Over Years

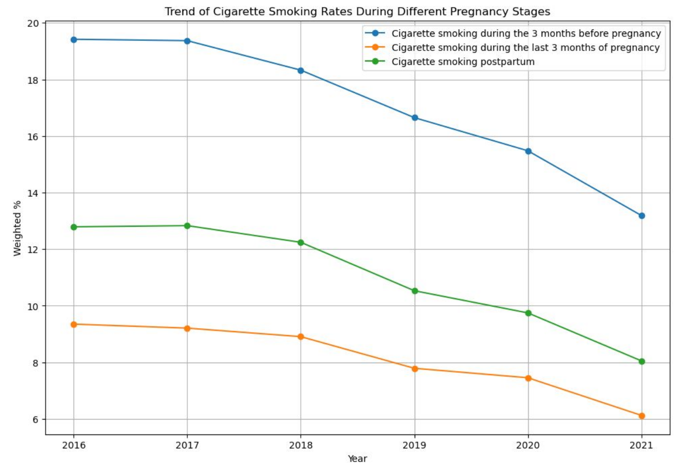

Fig. 7. Trend of Cigaratte Smoking

A. *Trend of Cigarette Smoking*

”Trend of Cigarette Smoking Rates During Different Pregnancy Stages” is an integral part of the experimental section, depicting the trends in cigarette smoking at three distinct stages: before pregnancy, during the last three months of pregnancy, and postpartum. This line graph shows a descending trend across all three stages from 2016 to 2021, with each stage represented by a different colored line for clarity. The blue line, indicating smoking rates during the three months before pregnancy, starts the highest but shows a consistent decline over the six years, suggesting a possible reduction in pre-pregnancy smoking habits. The green line, representing smoking during the last three months of pregnancy, also decreases but remains above the rates of postpartum smoking. This could reflect the effects of pregnancy-related health promotion or cessation efforts during prenatal care. The orange line, marking postpartum smoking rates, starts at the lowest and continues to decrease, possibly due to ongoing public health initiatives promoting the health of mothers and infants.

B. *Average Weighted Percentage of Multivitamin Use*

We further enriched our analysis by examining the geographical variations in health behaviors, particularly focusing on the average weighted percentage of multivitamin use among pregnant women from 2016 to 2021 across different U.S. states and territories. This horizontal bar graph presents a clear comparison of multivitamin use, highlighting significant variations from one region to another.

The graph categorizes states and territories by color, ranging from dark purple to yellow, with the length of each bar corresponding to the average weighted percentage of multivitamin use in that location. Puerto Rico, Louisiana, and Mississippi are at the top of the chart with the highest usage rates, indicated by the longest bars in dark purple. Conversely, states like New Hampshire, Massachusetts, and the District of Columbia are represented in yellow at the bottom, showcasing the lowest multivitamin use percentages.

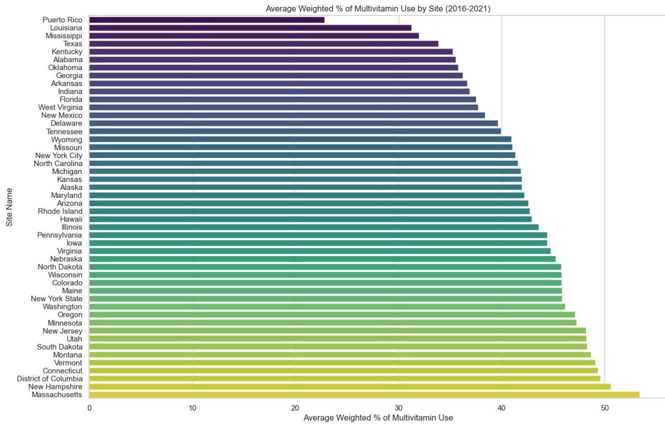

Fig. 8. Use of Multivitamin by State

This state-wise breakdown was instrumental in identifying patterns and outliers in prenatal care practices, providing a granular view of maternal health behaviors across the country. These findings prompted a more in-depth analysis of the social, economic, and policy-driven factors that could be contributing to the observed disparities, informing targeted public health interventions and resource allocation. The data visualization techniques employed were crucial for succinctly conveying these regional differences and facilitating a straightforward interpretation of a complex dataset.

C. Feature Selection by GMB Model

the Gradient Boosting Machine (GBM) model's feature importance analysis revealed critical insights into the determinants of maternal and child mortality rates. The model identified flu vaccination within 12 months before delivery as the most influential factor, suggesting a potentially protective role against mortality. State-specific factors emerged as significant, with New York and Texas showing notable impacts, indicating regional health disparities or policy effects. The analysis also highlighted the importance of insurance coverage, with both lower and upper confidence intervals of the uninsured rate showing a strong relationship with mortality rates. Early initiation of prenatal care was another key factor, underscoring its importance in maternal and child health. Lifestyle factors, such as hookah smoking, were associated with health outcomes, reflecting the influence of personal behaviors on health. The analysis also pointed to the complex role of pregnancy intentionality, with uncertainty around wanting pregnancy being a significant predictor, hinting at the psychological and social facets of health. Mental health factors, specifically postpartum depressive symptoms, were indicated as important, aligning with the growing recognition of mental health's impact on overall well-being. Surprisingly, the analysis suggested a negative association with heavy drinking before pregnancy, prompting further investigation to unravel this counterintuitive finding. Collectively, these results from the GBM model provide a multifaceted view of the variables affecting health outcomes, invaluable for informing public health strategies and interventions.

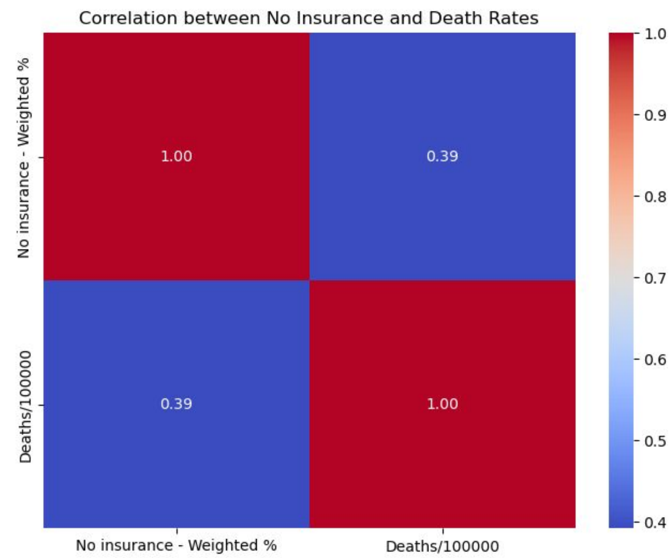

Fig. 9. Enter Caption

Factor	Weight
Had a flu shot in the 12 months before delivery - Weighted %	0.136197
Site Name_New York State	0.096896
Site Name_Texas	0.079479
No insurance - Lower 95% CI	0.067407
No insurance - Upper 95% CI	0.065457
Began prenatal care in 1st trimester - Weighted %	0.054576
Had a flu shot in the 12 months before delivery - Lower 95% CI	0.053055
No insurance - Unweighted Frequency	0.046556
Hookah use in the last 2 years - Weighted %	0.039227
Unwanted pregnancy - Unweighted Frequency	0.035086
Unsure whether wanted pregnancy - Unweighted Frequency	0.034439
Self-reported postpartum depressive symptoms - Lower 95% CI	0.023435
Self-reported depression before pregnancy - Unweighted Frequency	0.018435
Self-reported postpartum depressive symptoms - Unweighted Frequency	0.015902
Cigarette smoking before pregnancy - Weighted %	0.015539
Cigarette smoking postpartum - Lower 95% CI	0.014996
Hookah use in the last 2 years - Unweighted Frequency	0.014871
Underweight (BMI ¡ 18.5) - Unweighted Frequency	0.014799
Use of any postpartum contraception - Upper 95% CI	0.011062
Least effective contraceptive methods - Unweighted Frequency	0.010915

TABLE I
WEIGHTED FACTORS INFLUENCING MATERNAL MORTALITY

D. Correlation

We analyzed the correlation between insurance coverage and maternal mortality rates, depicted in the heatmap provided. The visualization shows a moderate positive correlation (0.39) between the lack of insurance and an increase in death rates per 100,000 individuals. This correlation coefficient suggests that as the percentage of uninsured individuals rises, there tends to be an increase in maternal mortality rates, though this relationship is not strong enough to suggest causation. The heatmap is an effective tool for quickly highlighting these relationships and guiding further investigation into how insurance coverage impacts maternal health outcomes.

VII. Future Scope

Moving forward, enhancing the GBM model with additional data could further improve predictions in maternal mortality. There's potential for real-time monitoring systems to leverage GBM insights for prompt interventions, and for these findings to be integrated with mHealth technologies, supporting personalized healthcare. Moreover, assessing policy impacts through the model's lens could guide effective healthcare strategies, while incorporating wider socioeconomic factors may sharpen our understanding of maternal mortality drivers.

VIII. Conclusion

Drawing upon the comprehensive exploration of maternal mortality rates (MMR), particularly within the United States, our project has illuminated the intricate tapestry of factors influencing this pressing health concern. The integration of a robust dataset from the CDC and the strategic application of the Gradient Boosting Machine (GBM) model has facilitated a nuanced understanding of the variables at play. With the GBM model's prowess in deciphering complex, non-linear data relationships, we established a mean squared error (MSE) of 177.49, affirming the model's predictive strength.

Our analysis did not stop at statistical modeling; we delved into multifaceted socio-behavioral dimensions. By examining the impact of lifestyle, insurance coverage, and mental health aspects on MMR, we were able to account for 34 of the variation in maternal death rates. This not only underscores the multifarious nature of the issue but also stresses the importance of behavioral and psychosocial elements in maternal health outcomes.

In conclusion, the insights gleaned from this study serve as a clarion call for a holistic approach to maternal health, one that transcends traditional medical perspectives. The data-driven evidence points towards the need for policy interventions that address the broader social determinants of health, advocate for comprehensive insurance coverage, and promote mental health support for expectant and new mothers. As the quest to reduce MMR continues, our research contributes a crucial piece to the complex puzzle, advocating for informed, inclusive, and proactive public health strategies.

References

[1] "Imagine a World,"www.gatesfoundation.org. https://www.gatesfoundation.org/goalkeepers/report/2023-report/

[2] "CDC WONDER," wonder.cdc.gov. https://wonder.cdc.gov/Welcome.html

[3] "Data Warehouse," UNICEF DATA. https://data.unicef.org/resources/data-explorer/unicef-f/?ag=UNICEFdf=GLO (accessed Dec. 15, 2023).

[4] World Health Organization, "Maternal Mortality," Who.int, Feb. 22, 2023. https://www.who.int/news-room/fact-sheets/detail/maternal-mortality.

[5] M. F. MacDorman, M. Thoma, E. Declercq, and E. A. Howell, "Causes contributing to the excess maternal mortality risk for women 35 and over, United States, 2016–2017," PLOS ONE, vol. 16, no. 6, p. e0253920, Jun. 2021, doi: https://doi.org/10.1371/journal.pone.0253920.

[6] S. Lisonkova et al., "Maternal age and severe maternal morbidity: A population-based retrospective cohort study," PLOS Medicine, vol. 14, no. 5, p. e1002307, May 2017, doi: https://doi.org/10.1371/journal.pmed.1002307.

[7] "Data export," WHO Data. https://platform.who.int/data/maternal-newborn-child-adolescent-ageing/data-export (accessed Dec. 15, 2023).

[8] "Query Data," UNICEF DATA. https://data.unicef.org/dv-index/?q=maten (accessed Dec. 15, 2023).

[9] "CDC WONDER," wonder.cdc.gov. https://wonder.cdc.gov/Welcome.html