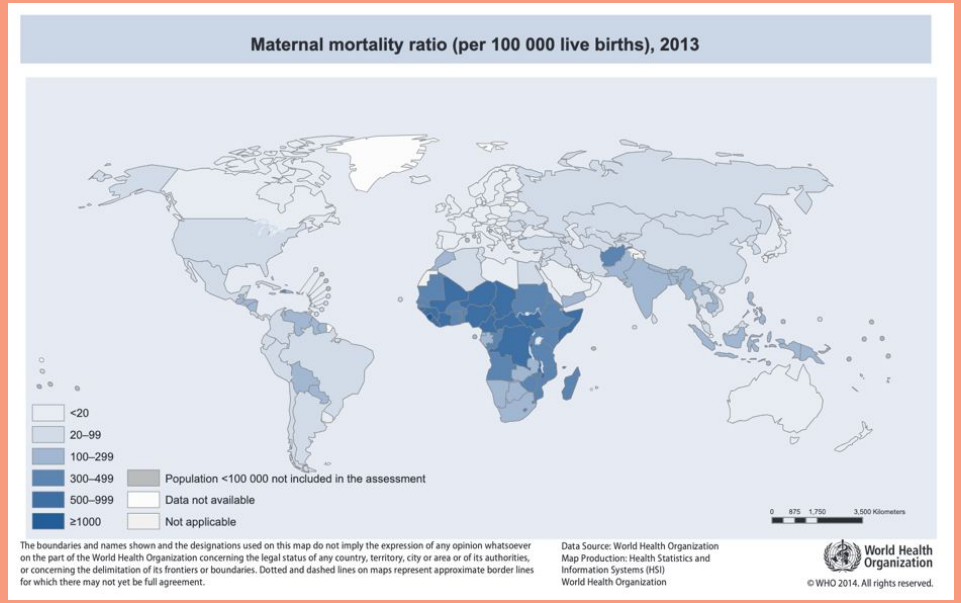
Analyzing US Maternal Mortality

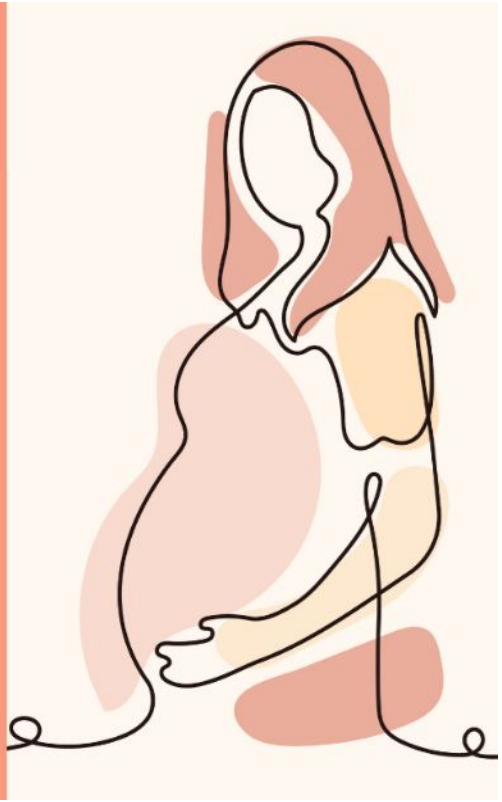
Team Members: David Clancy, Fei Li, Katherine Martin, Kristina Callaghan, Stephanie Egler, Solange Aveiro

Background and Motivation



Stakeholders:

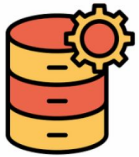
- Public health officials and professionals
- People who are pregnant, may become pregnant, or care about someone who is pregnant



Data Collection

?

What data did you use/collect? (How is that data suited to answer your research question)? How is that data structured?



The data used in this study include various demographic, socioeconomic, and medical factors related to birth outcomes.

The dataset is well-suited to answer the research question as it contains relevant features such as gestational age at birth, type of delivery, prenatal care initiation, and income brackets, which are all critical in assessing birth outcomes.

Data Sources

WONDER Database for state-level statistics of maternal deaths, births, and various natality information.

Division of Nutrition, Physical Activity, and Obesity database for state-level health-related statistics.

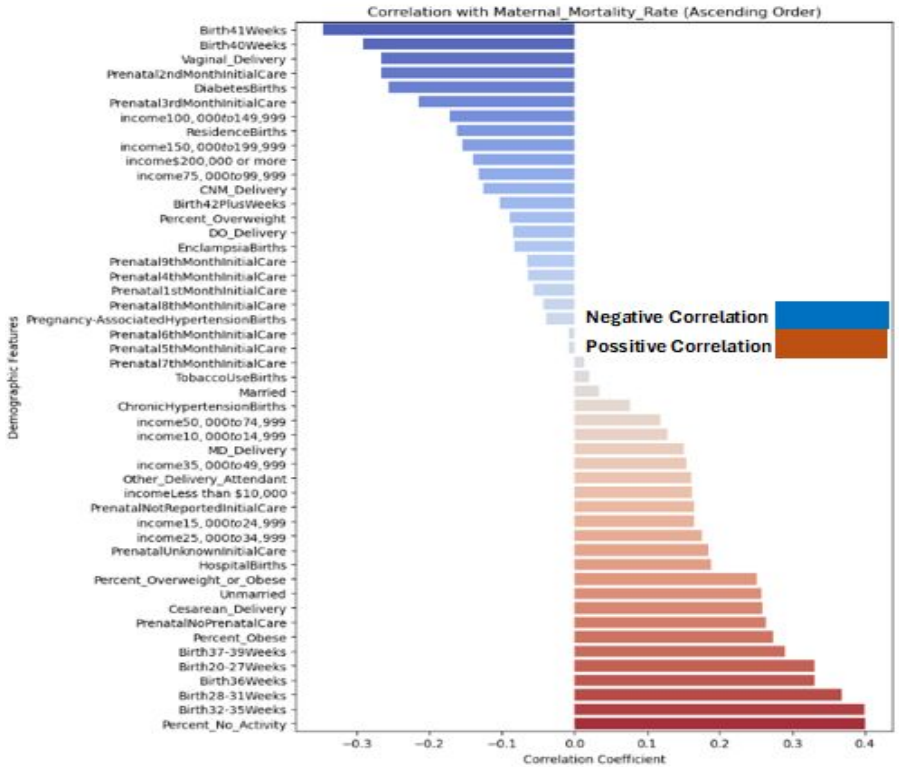
American Community Survey for state-level information on income distributions.

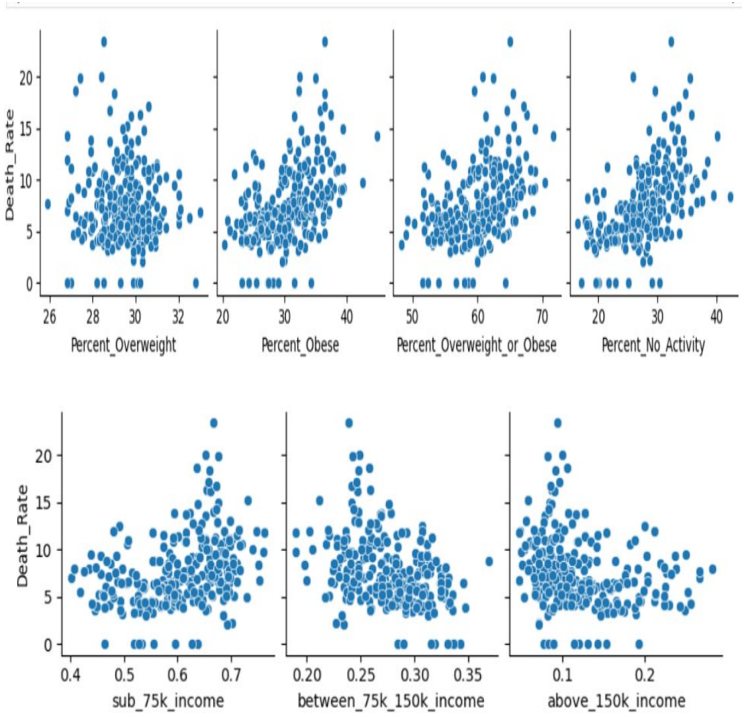
Key Predictors of Maternal Mortality

Features:

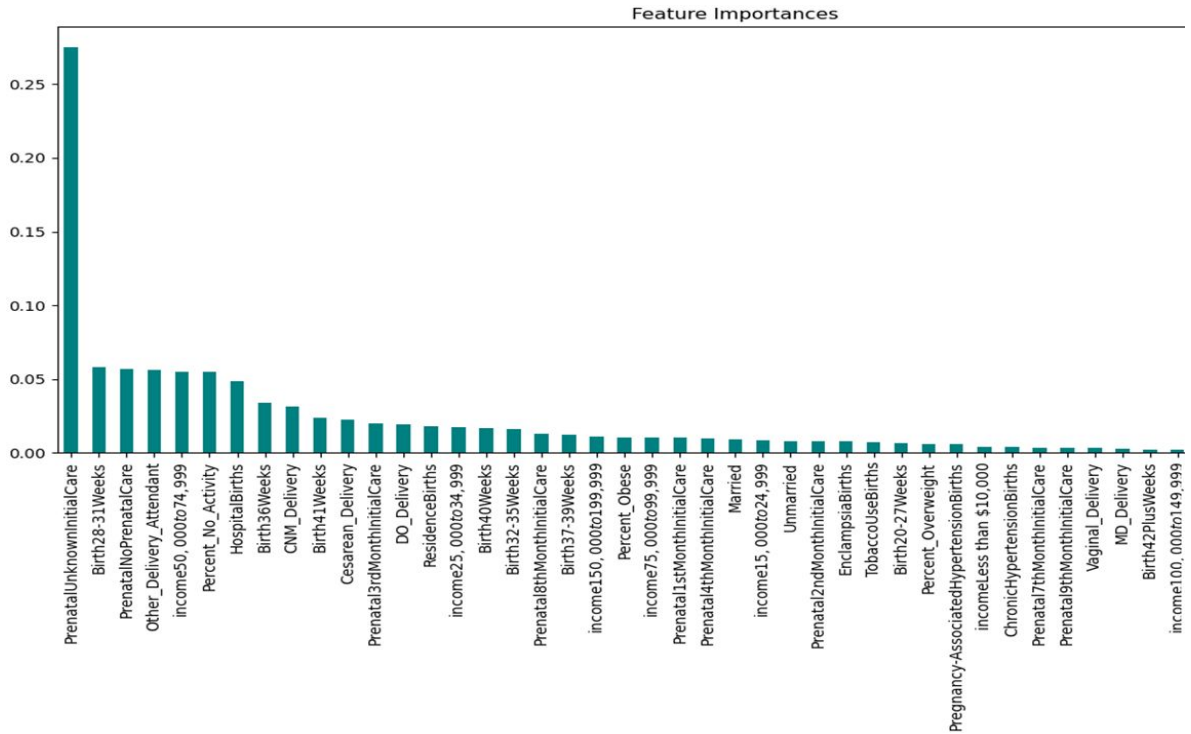
- Income brackets
- Health Indicators
- Healthcare Access
- Delivery and Complications

Correlations between target and features





Feature Importance Analysis




```
Cross-validated RMSE score: 104.00
Cross-validated R^2 score: 0.21
Test Set Mean Squared Error: 87.61
Test Set Root Mean Squared Error: 9.36
Test Set R^2: 0.52
Feature Importances:
PrenatalUnknownInitialCare                0.275123
Birth28-31Weeks                           0.057761
PrenatalNoPrenatalCare                    0.056728
Other_Delivery_Attendant                  0.056203
income$50,000 to $74,999                  0.054894
Percent_No_Activity                       0.054773
```

Modeling Approach

Start

Create a pipeline to train on different scalings and regression models

Use GridSearchCV and a 5-fold cross-validation to determine best combination of scalings, models, and model parameters

BEST

Rank results by the lowest average mean squared error

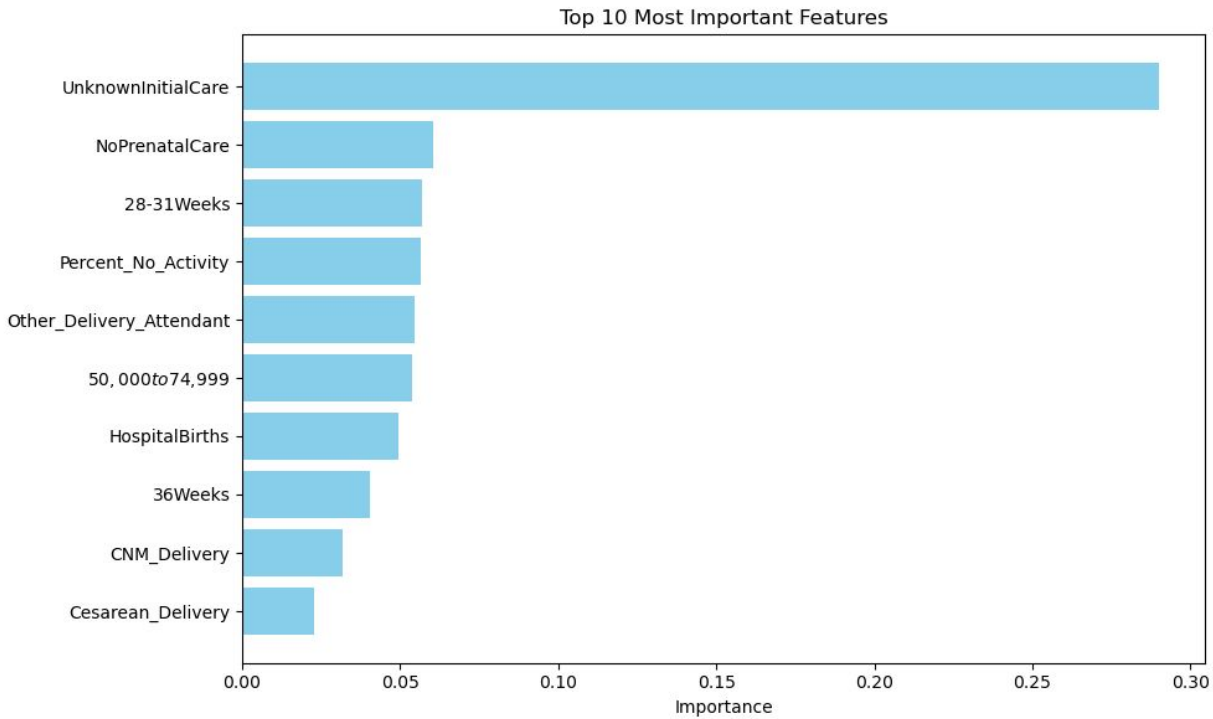
WORST

Gradient Boosting Regressor	KNeighbors Regressor	Random Forest Regressor	Ridge Regression	Support Vector Regression	Lasso Regression	Elastic Net Regression	Linear Regression	Decision Tree Regressor	Baseline Model

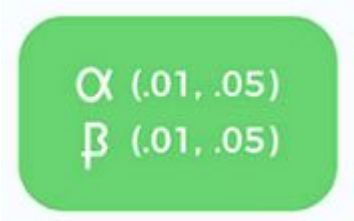
Best Model Performance

Gradient Boosting Regressor MSE: 96.8

Baseline Model MSE: 182.0



Approach Strengths and Weaknesses



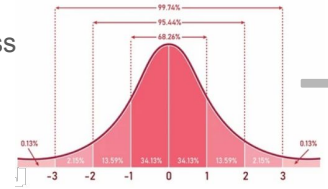
Was your approach accurate? What were its strengths / weaknesses?

The best model was found using gradient boosting

With an R^2 value of 0.46 and a MSE of 97

Weaknesses

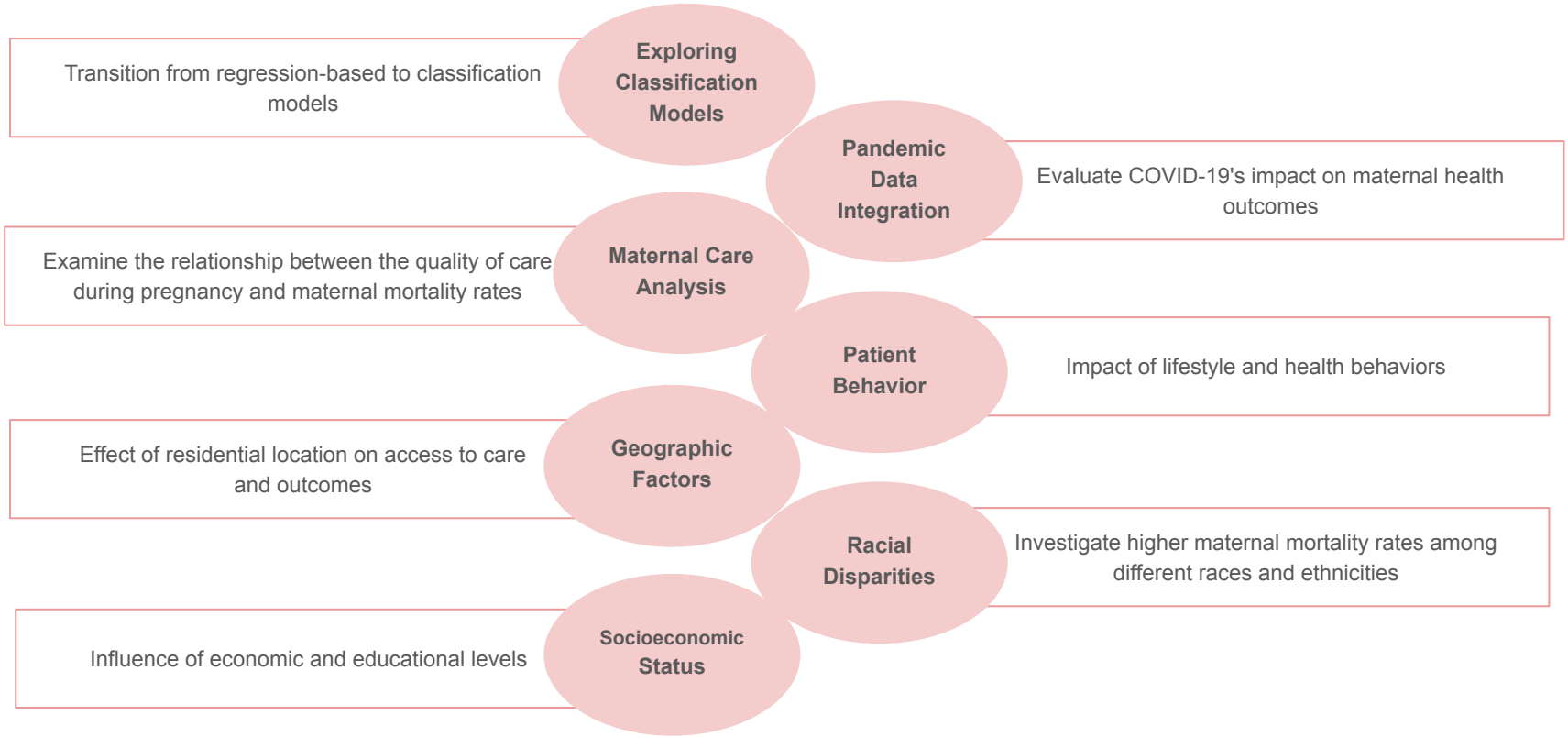
The greatest weakness for our model(s) was limited data.



Strengths

The features that this model determined as important correspond with what we expected.

Next Steps


Exploring Classification Models
Transition from regression-based to classification models
Pandemic Data Integration
Evaluate COVID-19's impact on maternal health outcomes
Maternal Care Analysis
Examine the relationship between the quality of care during pregnancy and maternal mortality rates
Patient Behavior
Impact of lifestyle and health behaviors
Geographic Factors
Effect of residential location on access to care and outcomes
Racial Disparities
Investigate higher maternal mortality rates among different races and ethnicities
Socioeconomic Status
Influence of economic and educational levels

Thank you!