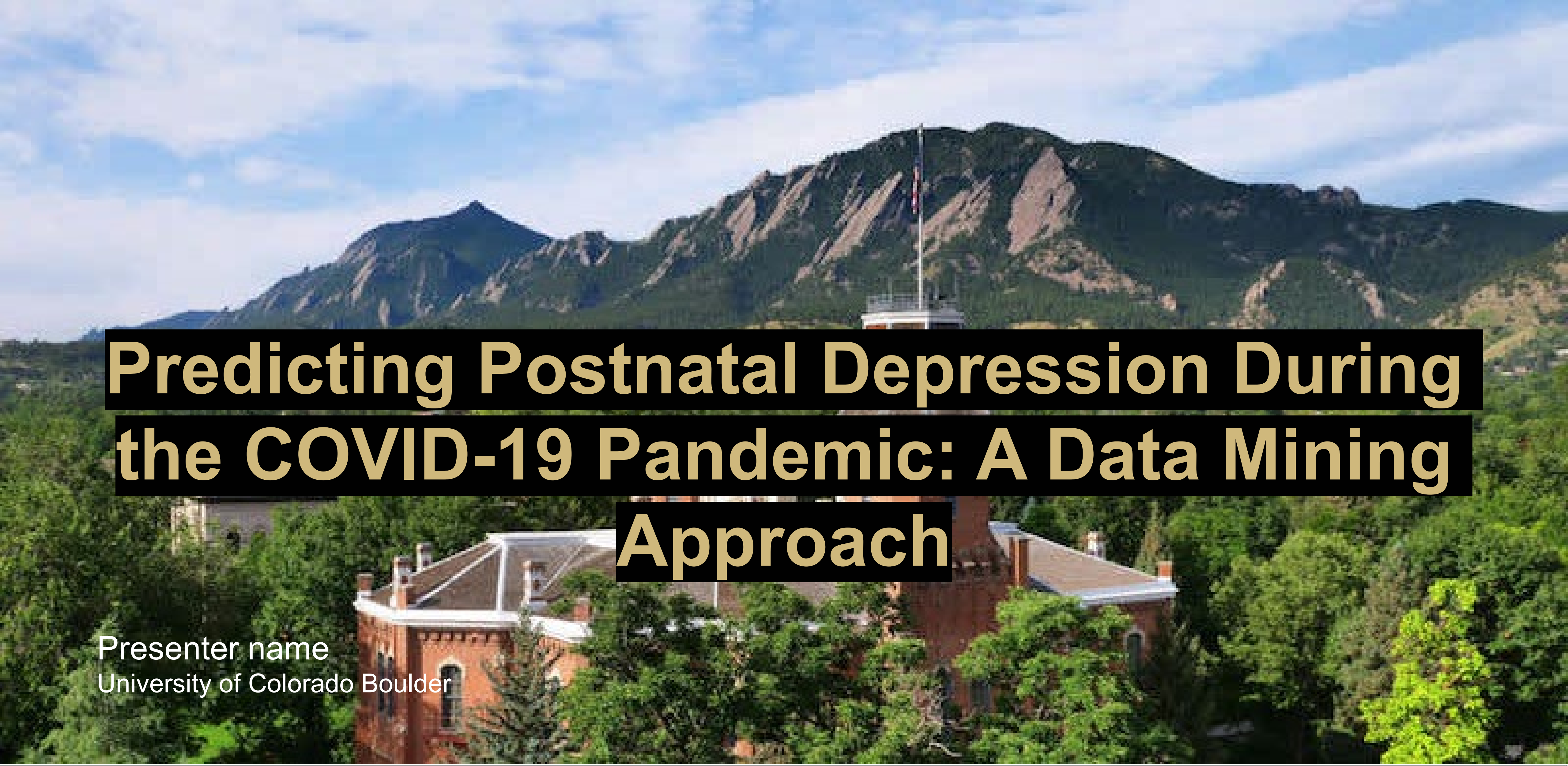


Predicting Postnatal Depression During the COVID-19 Pandemic: A Data Mining Approach

Presenter name
University of Colorado Boulder

DATA SCIENCE
UNIVERSITY OF COLORADO **BOULDER**

Abstract

This project aims to develop a predictive model for identifying individuals at high risk of postnatal depression using data collected during the COVID-19 pandemic, highlighting the use of data mining techniques and the potential impact on healthcare interventions.

Sources of Data

- The Pregnancy During the COVID-19 Pandemic (PdP) study was designed to investigate the associations between exposure to objective hardship caused by the pandemic and psychological distress in pregnant individuals, and developmental outcomes in their offspring.
- More than 11,000 responses from social media beginning on April 5, 2020
- The PdP study comprises a prospective longitudinal cohort of individuals who were pregnant at enrollment, with repeated follow-ups during pregnancy and the postpartum period. Participants were eligible if they were pregnant, ≥17 years old, at ≤35 weeks of gestation at study enrollment, living in Canada

Sources of Data 2

- Maternal_Age
- Household_Income
- Maternal_Education
- EPDS: Edinburgh Postnatal Depression Scale
- PROMIS_Anxiety: Higher scores indicating greater severity of anxiety.
- GAbirth: Gestational age at birth (in weeks)
- Delivery_Date
- Birth_Length
- Birth_Weight
- Delivery_Mode
- NICU_stay
- Threaten_Life: How much do (did) you think your life is (was) in danger during the pandemic? (0-100)
- Threaten_Baby_Danger: How much do (did) you think your unborn baby's life is (was) in danger during the pandemic? (0-100)
- Threaten_Baby_Harm: How much are you worried that exposure to the COVID-19 virus will harm your unborn baby? (0-100)

Introduction

- The COVID-19 pandemic has heightened anxiety and stress among pregnant individuals, leading to increased mental health challenges due to social isolation and healthcare disruptions.
- Early detection of postnatal depression is vital for the health of both mothers and infants. Timely intervention can prevent long-term psychological issues and improve overall outcomes.
- Objectives:
  - Develop a predictive model for postnatal depression.
  - Identify key predictors from demographic, mental health, and perceived threat data.
  - Support healthcare interventions by providing actionable insights.
  - Enhance maternal and infant health through early detection and support.

Problem Statement

- Increased Stress During Pregnancy Due to COVID-19:
  - The pandemic has significantly elevated stress levels in pregnant individuals, contributing to mental health challenges.
- Goal of Predicting Postnatal Depression:
  - Develop a model to predict postnatal depression using the Edinburgh Postnatal Depression Scale (EPDS), incorporating pandemic-specific stress factors.
- Importance of Early Identification:
  - Early identification of at-risk individuals is crucial for timely intervention, reducing the risk of long-term mental health issues for both mothers and infants.

Proposed Work

- Data Cleaning: Handling missing values and converting categorical variables.
- EDA: Visualizing data relationships and identifying correlations.
- Feature Engineering: Selecting relevant features.
- Modeling: Training and evaluating regression models.
- Evaluation: Using MSE and R^2 metrics.
- Feature Importance: Identifying key features using the Random Forest model.

Evaluation Plan

- Metrics for assessing model performance: MSE and R^2.
- Criteria for success: low MSE, high R^2, and actionable insights.
- Importance of feature importance analysis for refining the model.

Timeline

- Week 3: Data Cleaning and Preparation
- Week 4: Exploratory Data Analysis and Feature Engineering
- Week 5: Model Training and Initial Evaluation
- Week 6: Model Refinement and Feature Importance Analysis
- Week 7: Final Evaluation and Report Preparation

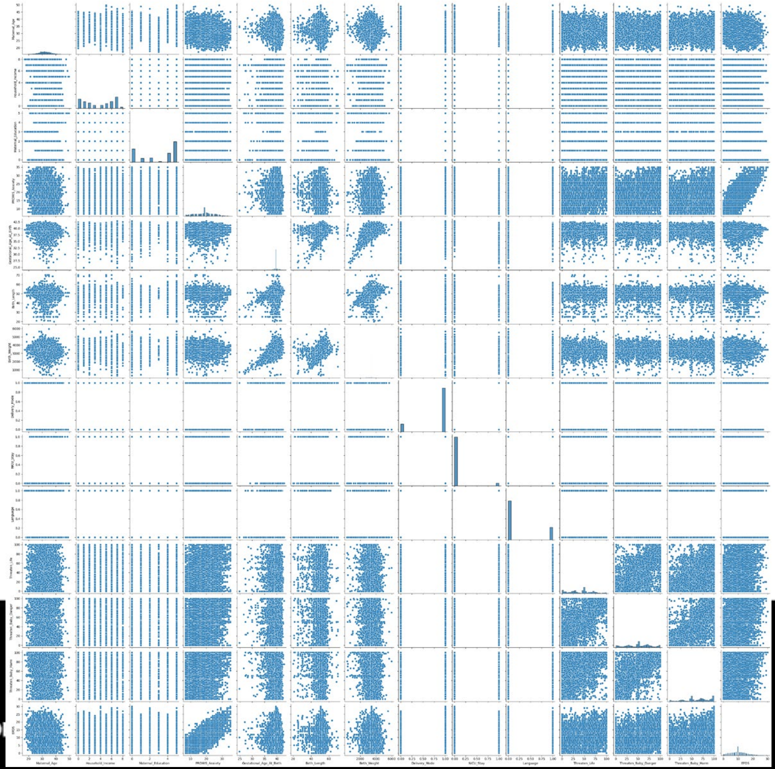
Process 1

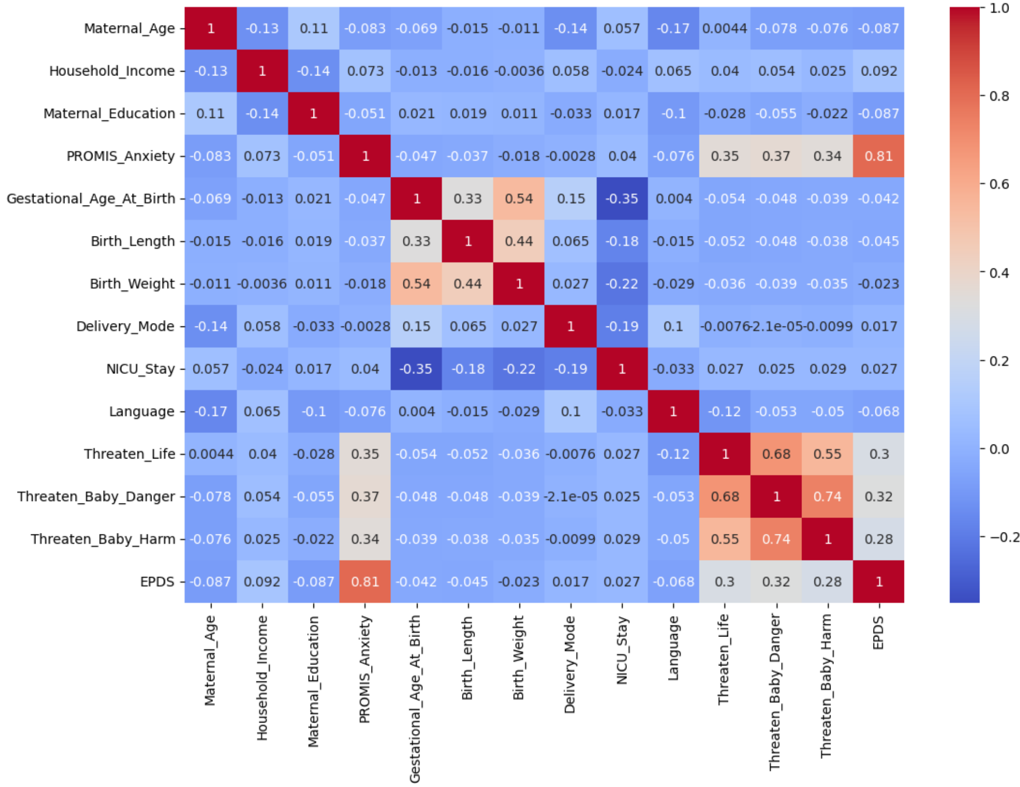
- Download data from OSF and save as CSV
- Cleaned data
  - Replace NA values
    - Numerical
    - Categorical
  - Convert objects to floats

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 10772 entries, 0 to 10771
Data columns (total 14 columns):
 #   Column                   Non-Null Count  Dtype  
---  ------                   --------------  -----  
 0   Maternal_Age             10661 non-null  float64
 1   Household_Income         10772 non-null  int8   
 2   Maternal_Education       10772 non-null  int8   
 3   PROMIS_Anxiety           10772 non-null  float64
 4   Gestational_Age_At_Birth 10772 non-null  float64
 5   Birth_Length             10772 non-null  float64
 6   Birth_Weight             10772 non-null  float64
 7   Delivery_Mode            10772 non-null  int8   
 8   NICU_Stay                10772 non-null  int8   
 9   Language                 10772 non-null  int8   
 10  Threaten_Life            9876 non-null   float64
 11  Threaten_Baby_Danger     9868 non-null   float64
 12  Threaten_Baby_Harm       9880 non-null   float64
 13  EPDS                     10772 non-null  float64
dtypes: float64(9), int8(5)
memory usage: 810.1 KB
```

Process 2

- Run pairplot and heatmap to identify relationships between columns
- Identify key features to find relationships with EPDS
  - 'Maternal_Age', 'Household_Income', 'Maternal_Education', 'PROMIS_Anxiety', 'Threaten_Life', 'Threaten_Baby_Danger', 'Threaten_Baby_Harm'

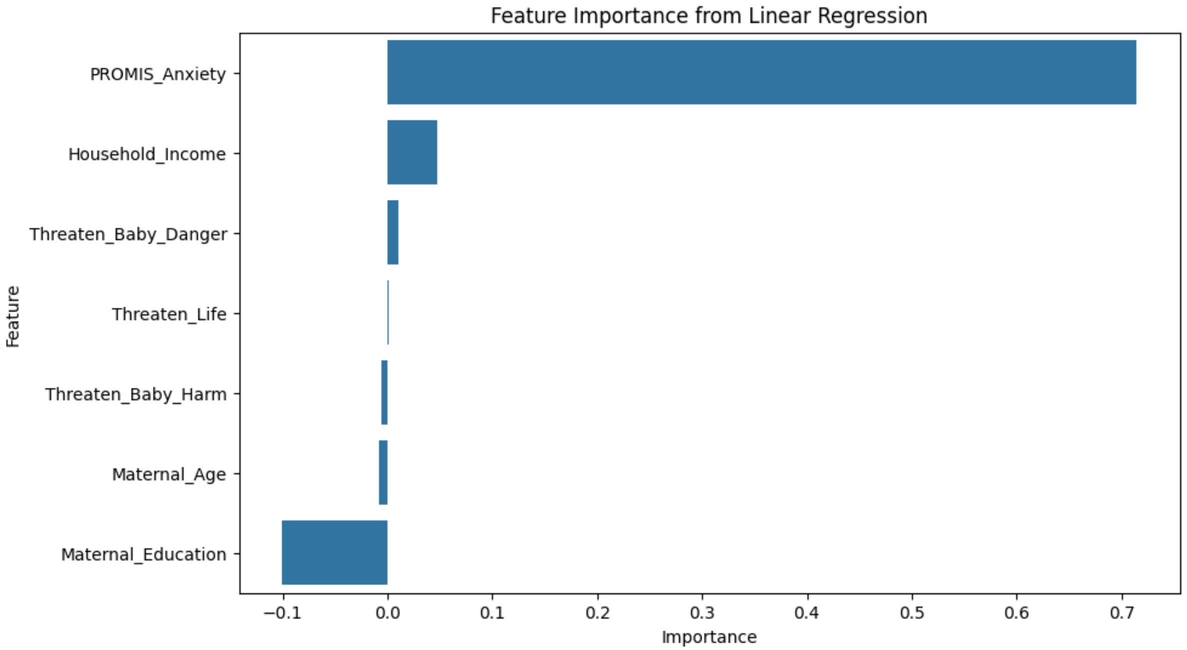




Process 3

- Run linear regression using EPDS as response

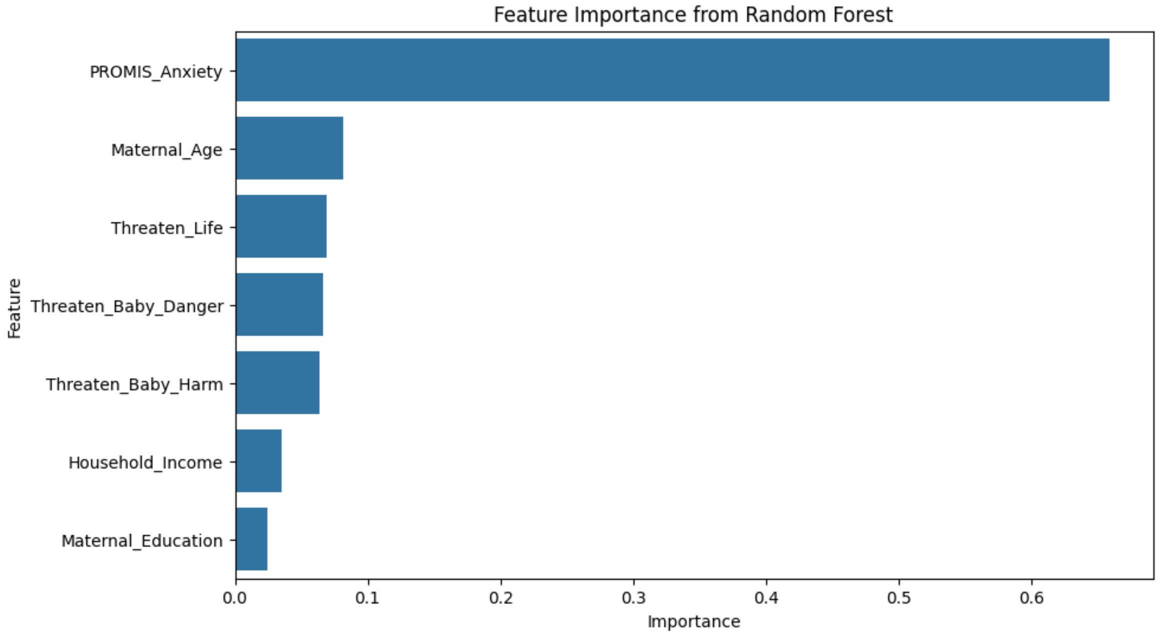
- MSE: 8.986360817947201, R2: 0.6620131234030475
- Feature Importance
  - Maternal_Age -0.008575
  - Household_Income 0.047906
  - Maternal_Education -0.101161
  - PROMIS_Anxiety 0.714697
  - Threaten_Life 0.000784
  - Threaten_Baby_Danger 0.010607
  - Threaten_Baby_Harm -0.006049



Process 4

- Run Random Forest using EPDS as response

- Random Forest Regressor - MSE: 9.453041111111111, R2: 0.64446076624187
- Feature Importance
  - PROMIS_Anxiety   0.658800
  - Maternal_Age   0.081931
  - Threaten_Life   0.069322
  - Threaten_Baby_Danger   0.066066
  - Threaten_Baby_Harm   0.064104
  - Household_Income   0.034901
  - Maternal_Education   0.024877



Conclusion

- The linear model performed better than the random forest
- The key indicator of postnatal anxiety is prenatal anxiety

Future Work

- Would love updated data on how the child is performing
- More demographic data would be great to see if anything else can impact postpartum anxiety levels
- Would love to compare to mothers before and after COVID

References

- Gao, W., Jalal, Z., Taylor, B. K., Qian, H., Reichert, A. R., & Blank, P. R. (2023). The impact of COVID-19 pandemic on mental health in pregnant individuals. The Lancet Regional Health – Europe, 24. https://doi.org/10.1016/j.lanepe.2023.100473

- Huang, Y., Alvernaz, S., Kim, S. J., Maki, P., Dai, Y., & Peñalver Bernabé, B. (2023). Predicting prenatal depression and assessing model bias using machine learning models. medRxiv. https://doi.org/10.1101/2023.07.17.23292587