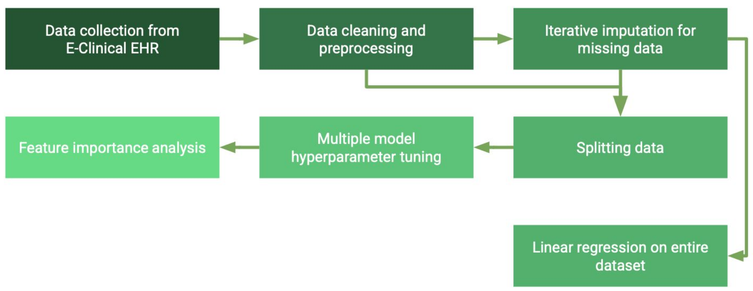
Assessing Patterns of Pediatric Obesity in Patients from Providence, RI with Machine Learning

Roshan S. Parikh
Data Science Institute, Brown University

Overview

Pediatric obesity is a major public health issue in the US. This project aimed to determine how the manifestation of obesity varies for children of different birth weight percentiles (per the Fenton scale).

Project Design



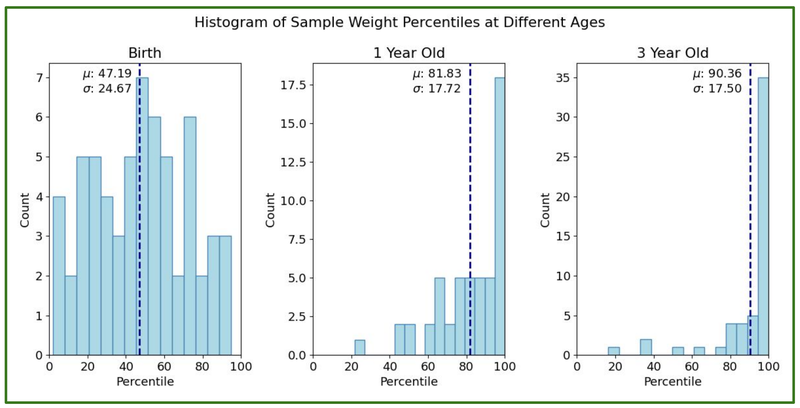
Data

Overview

- All children included in the study had a diagnosis of obesity by age 10.
- 2 target variables (birth weight percentile and a categorization of this percentile) and 10 features were collected. After preprocessing, there were 24 features.

Challenges

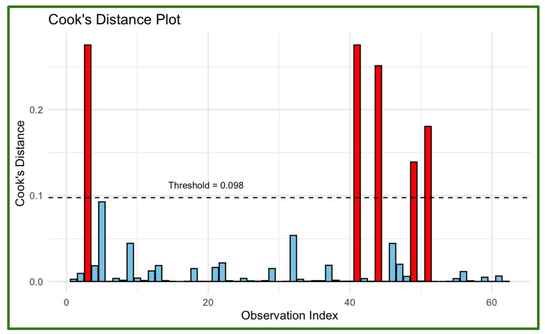
- One feature, age of onset of obesity, was anticipated to be unreliable, since the criteria of diagnosis are not consistent across providers.
- All features except sex and gestational age had missing features.



Statistical Methods

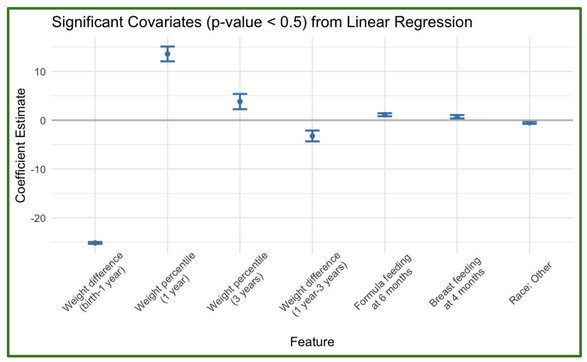
Learning

- A linear model was trained on the full dataset after iterative imputing to determine patterns in the data set. Outliers were removed before fitting the model again.



Results

- Residual standard error: 1.218 on 38 degrees of freedom
- Multiple R-squared: 0.9984, Adjusted R-squared: 0.9976
- F-statistic: 1303 on 18 and 38 DF, p-value: < 2.2e-16



Best Models

Machine Learning

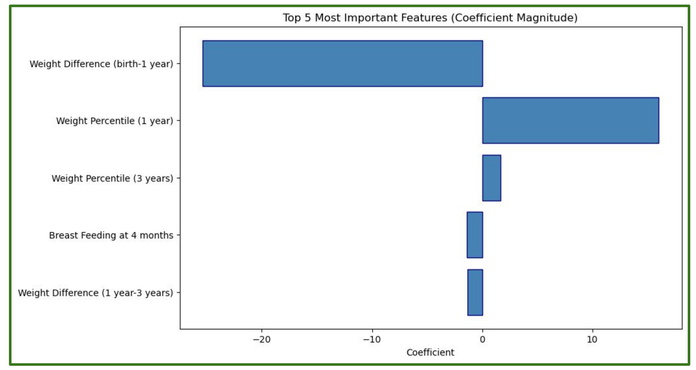
- Data was split into a 80-20 train/val-test split before k-fold validation (4 folds).
- Multiple models were trained with hyperparameter tuning.

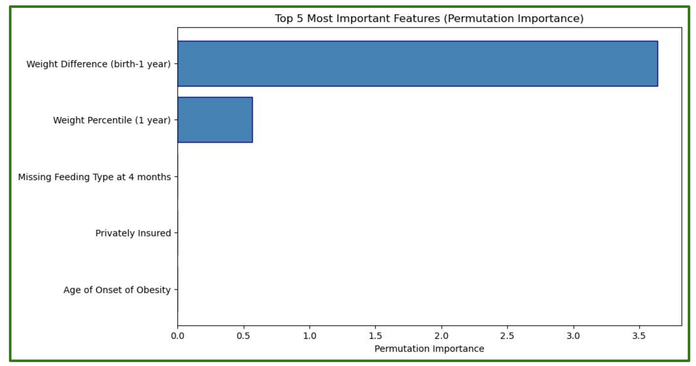
Model	Best RMSE	Improvement from Baseline
XGB	7.58%	66.3%
Linear Reg.	3.09%	86.2%
Ridge Reg.	3.01%	86.6%
Lasso Reg.	2.91%	87.0%
SVR Linear	1.85%	91.8%
SVR RBF	8.34%	62.9%
SVR Poly.	15.93%	29.1%

Feature Importance

Model

- Feature importance was assessed from the model that had the highest improvement of RMSE from baseline on the validation set. This was the **support vector regressor** with a **linear kernel**.





Conclusion

Discussion

- Children who were born with a lower birth weight percentile tended to have a larger-than-normal increase in weight in their first year.
- However, children who were born with a larger birth weight still had greater weights than those with a smaller birth weight at years 1 and 3.

Future Directions

- Collecting additional data from Sunshine Pediatrics to form more stable models.
- Performing this same pipeline as a classification task. That is, using the second target variable.
- Expanding the number of features during data collection.

Data Source: Data came from patient records at Sunshine Pediatrics in Providence, RI.