Elevance Health

Logan Portscheller | Nathan Wu | Taran Selvan | Jason Lyu | Akki Poojari | Olayemi Sanusi
Sarah Harman | Katrina Rose | Gibson Adema | Taylor Holdridge | Grace Enright

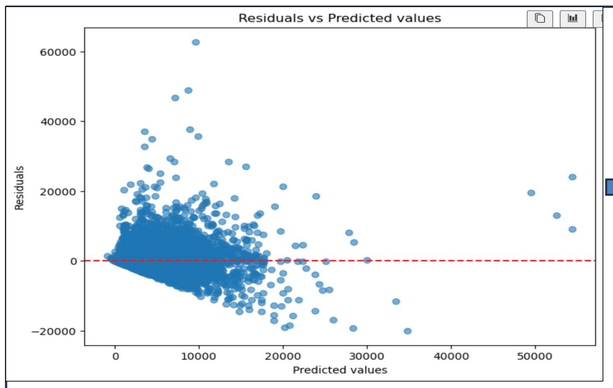
PURDUE UNIVERSITY | The Data Mine

Background

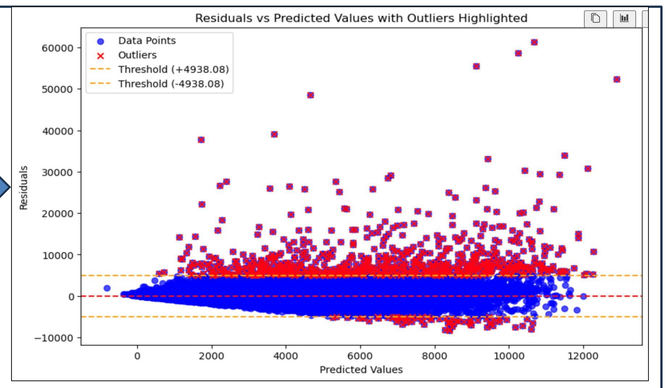
- **Elevance Health** is a health insurance company focused on improving health outcomes through innovative, technology-driven solutions and personalized care.
- **Objective:** Develop a process to identify provider billing changes.
- **Goals:** Utilize machine learning techniques to predict expected costs and identify key drivers of cost changes.

Error/Accuracy Metrics	Lasso	Ridge
Mean Absolute Error (MAE)	1,089.00	974.61
Root Mean Squared Error (RMSE)	1,965.94	1,800.24
R^2	0.48	0.56

(1) During our early stages we tested different regression and machine learning models to determine which model performed best on the data.

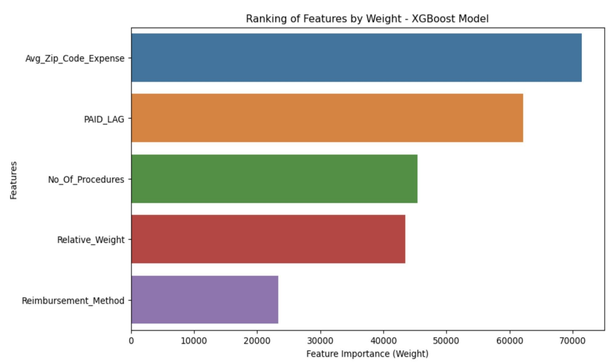


(2) This graph shows residuals before outlier handling. Residuals are widely scattered, with extreme outliers distorting model performance and reducing prediction accuracy.



(3) This graph shows residuals after outlier are handled and managed, leading to tighter residual distribution and improved model accuracy.

(4) To reduce the Mean Absolute Error and refine the model, detailed feature importance graphs were generated to identify the most influential variables in the predictions. This also helped determine which less significant variables could be removed to enhance overall performance and efficiency.



RESEARCH METHODOLOGY

Data Preprocessing → Machine Learning Model Training → Cost Predictions → Model Analysis

- Claims dataset provided by Elevance had information about emergency room visits
  - Dataset was in tabular format and had features such as severity and number of procedures of the visit. As well as the average allowed amount by zip code.
- Developed Predictive Models using:
  - Random Forest
  - XG Boost
  - Linear Regression
- Optimized Model Performance by handling outliers and fine-tuning hyperparameters for better predictions
- Enhanced Model Robustness using a regulated dataset, leading to reduced Mean Absolute Error (MAE)

RESULTS

- Random Forest
  - *Starting Mean Absolute Error (MAE)*: 623
  - *Changes Made*: Created new feature to normalize the average allowed amount per zip code. Found outliers with bounds based on IQR and factored them out for training, while still considering them for testing.
  - *Final Mean Absolute Error (MAE)*: 501
  - R^2: 0.785
- XGBoost:
  - Starting *Mean Absolute Error (MAE)*: 657
  - *Changes Made*: Capped outliers in the data using the Inter Quartile Range (IQR) method. Combined low-predictive features into clustering features for better performance.
  - *Final Mean Absolute Error (MAE)*: 594
  - R^2: 0.708

Conclusions

- **Key Costs:** Implemented both Random Forest and XGBoost algorithms to predict expected costs and identify key cost drivers.
- **Feature Engineering**: Implemented techniques such as pipelines, target encoding, and feature creation.
- **Outlier Management:** By observing outliers in the dataset that represent realistic data points. we carefully handled them to build a model that performs well for both outliers and non-outlier data.

Future Goals

- Continue to implement more advanced machine learning models such as:
  - Neural Networks
  - Support Vector Regressors
- Use results from machine learning techniques to determine major drivers of changes in billing practices and costs.
- Compare healthcare providers to identify those that stand out in terms of billing patterns and cost drivers.

References

Thank you to our incredible TAs, *Arnav* and *Katelyn*, for their support and guidance, and to our dedicated mentor, *Dylan*, for always being there for us!