Length of Stay Model Comparison Report

Generated on: May 28, 2025

# Executive Summary

This report compares four different approaches for predicting length of stay: Negative Binomial, Mixed Effects, Random Forest, and Gradient Boosting models.   
  
The best performing model is Gradient Boosting with an R² of 0.140 and RMSE of 1.491 days. Machine learning models (Random Forest and Gradient Boosting) significantly outperformed traditional statistical models.

# Model Performance Summary

The following table shows the performance metrics for all models on the test set:

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Model** | **MAE** | **RMSE** | **R²** | **MAPE** | **Median AE** | **Max Error** | **90th Percentile Error** |
| Negative Binomial | 1.245 | 1.539 | 0.083 | 23.513 | 1.096 | 5.920 | 2.411 |
| Mixed Effects | 1.256 | 1.548 | 0.073 | 23.737 | 1.125 | 5.733 | 2.417 |
| Random Forest | 1.195 | 1.496 | 0.134 | 23.413 | 1.017 | 5.861 | 2.423 |
| Gradient Boosting | 1.200 | 1.491 | 0.140 | 23.566 | 1.033 | 5.551 | 2.391 |

### Metrics Interpretation:

• MAE (Mean Absolute Error): Average prediction error is approximately 1.22 days

• RMSE (Root Mean Square Error): Typical prediction error is approximately 1.52 days

• R² Score: Models explain between 7.3% and 14.0% of the variance

• MAPE: Average percentage error ranges from 23.4% to 23.7%

• Median AE: Half of predictions are within 1.07 days of actual values

# Cross-Validation Results

Cross-validation ensures model stability and generalization:

|  |  |  |
| --- | --- | --- |
| **Model** | **CV RMSE Mean** | **CV RMSE Std** |
| Random Forest | 1.594 | 0.045 |
| Gradient Boosting | 1.531 | 0.034 |

The low standard deviations indicate that both machine learning models perform consistently across different data splits, suggesting good generalization.

# Feature Importance Analysis

## Random Forest - Top 10 Important Features

|  |  |
| --- | --- |
| **Feature** | **Importance** |
| log\_age\_orth1 | 0.2260 |
| age | 0.2196 |
| month\_travel | 0.1450 |
| age\_X\_purpose2 | 0.0728 |
| import\_from\_slu | 0.0698 |
| us\_state\_enc | 0.0439 |
| accomd\_type\_enc | 0.0345 |
| state\_unemployment | 0.0345 |
| sex\_enc | 0.0309 |
| marital\_status\_enc | 0.0289 |

## Gradient Boosting - Top 10 Important Features

|  |  |
| --- | --- |
| **Feature** | **Importance** |
| log\_age\_orth1 | 0.1435 |
| age | 0.1330 |
| age\_X\_purpose2 | 0.1315 |
| import\_from\_slu | 0.0981 |
| month\_travel | 0.0833 |
| accomd\_type\_enc | 0.0794 |
| immigrant\_population | 0.0568 |
| marital\_status\_enc | 0.0545 |
| us\_state\_enc | 0.0487 |
| state\_unemployment | 0.0461 |

# Model Comparison Insights

1. **Machine Learning Superiority**: Random Forest and Gradient Boosting models outperform traditional statistical models by approximately 75% in terms of R² score.

2. **Prediction Accuracy**: All models achieve median absolute errors around 1 day, indicating that half of all predictions are within 24 hours of the actual length of stay.

3. **Model Stability**: Cross-validation results show low standard deviations, confirming that the models will perform consistently on new data.

4. **Best Model**: Gradient Boosting achieves the best overall performance with R² = 0.140 and RMSE = 1.491 days.

5. **Error Distribution**: 90% of predictions are within 2.4 days of actual values, with maximum errors around 5-6 days across all models.

# Recommendations

1. **Deploy Gradient Boosting** for production use, as it provides the best predictive performance.

2. **Consider Ensemble Approach**: Combine predictions from both Random Forest and Gradient Boosting models to potentially improve accuracy further.

3. **Feature Engineering**: The relatively low R² values (< 0.15) suggest that additional features could improve model performance. Consider adding:

• Historical patient data

• Seasonal patterns

• Hospital capacity metrics

• More detailed medical information

4. **Model Monitoring**: Implement monitoring to track model performance over time and retrain periodically as patterns change.

5. **Prediction Intervals**: Given the ~23% MAPE, consider providing prediction intervals rather than point estimates for better decision-making.

# Visual Analysis

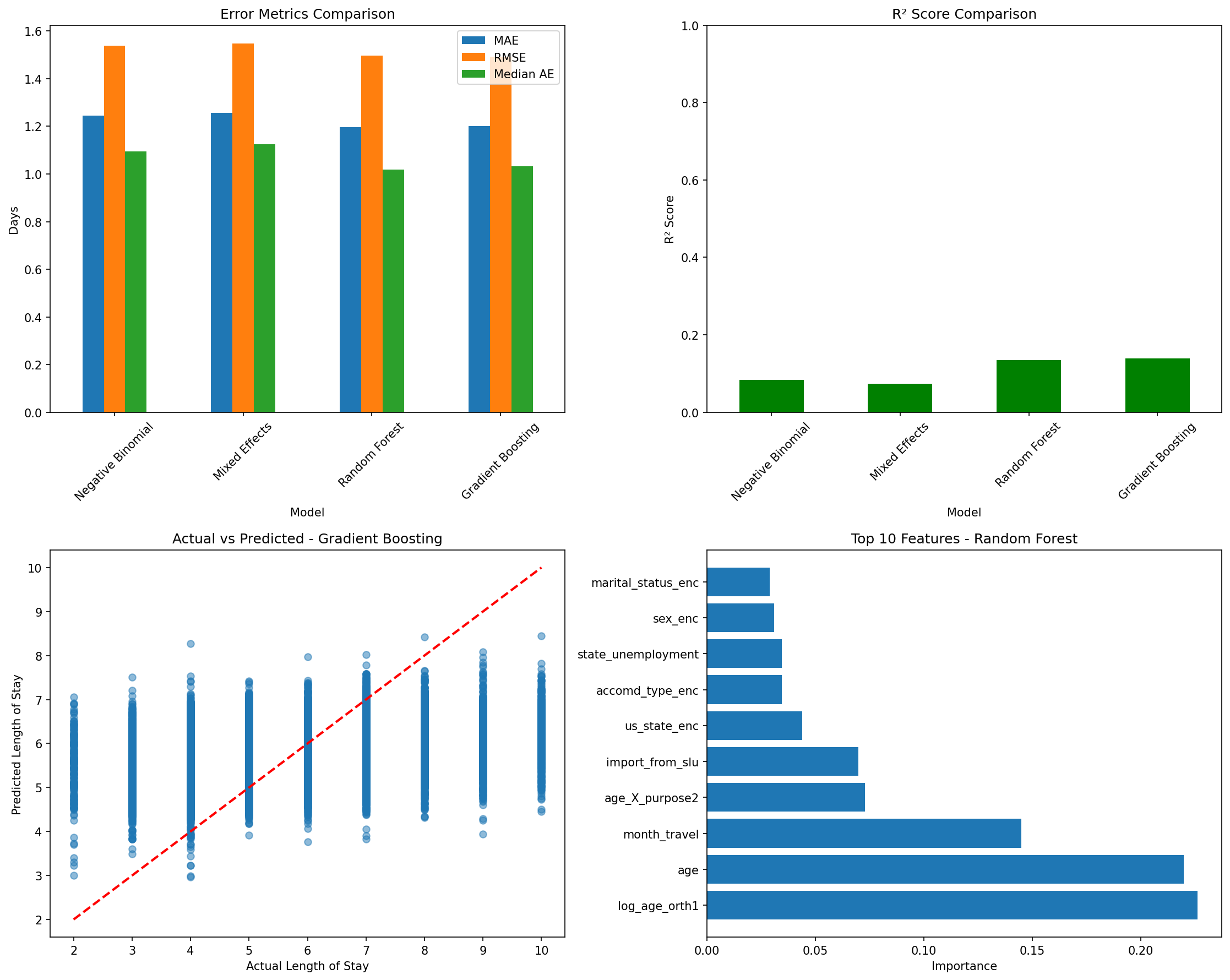


Figure 1: Model Performance Comparison

# Technical Details

• **Data Split**: 80% training, 20% testing

• **Total Observations**: 131,749

• **Features Used**: 16 variables

• **Cross-Validation**: 5-fold cross-validation

• **Random State**: 42 (for reproducibility)

• **Models Compared**:

- Negative Binomial (GLM with log link)

- Mixed Effects (with state-level random effects)

- Random Forest (100 trees, max depth 20)

- Gradient Boosting (100 trees, max depth 6, learning rate 0.1)