Detailed Report of Attempts to Predict Patient Outcomes (ICU vs. General Ward)

Our analysis journey began with the goal of predicting meaningful patient outcomes from a healthcare dataset related to burn injuries. Here's a detailed walkthrough of each approach we tried, the challenges we faced, and how we arrived at the final best-suited model for predicting ICU vs. General Ward with 86% accuracy.

Initial Goal: Predicting Hospitalization (Admission)

We started with the goal of predicting whether a patient would require hospitalization. The primary challenge here was identifying the most important factors for deciding whether a patient would need to be admitted.

Approach 1: Predicting Admission

Selected Features:

We chose key features like: Ventilator Days (daysvent): Indicates whether a patient required ventilator support. Total Burn Surface Area (tbsabrn): Measures the severity of the burn. Age (age): Age can affect a patient's vulnerability. Challenges:

Upon running the initial Random Forest model, we found that 99.89% of patients were classified as "admitted", which suggested that the dataset was heavily skewed or that the model was simply labeling everyone as admitted. Outcome:

This led us to rethink our approach since almost every patient was predicted to require hospitalization, providing little room for meaningful prediction. Refocusing on Predicting ICU vs. General Ward Admission After realizing that predicting hospitalization wasn't providing actionable insights, we shifted our focus to predicting whether a patient required ICU care or could be treated in the General Ward. This approach promised more nuanced insights.

Approach 2: Predicting ICU vs. General Ward

Chosen Features:

We initially selected features that directly related to the severity of the injury or the level of care required: Ventilator Days (daysvent) Total Burn Surface Area (tbsabrn) Rehabilitation Days (rhb_days) Age (age) Challenges:

We achieved near-perfect accuracy (99.64%) in the initial Random Forest model, but this raised concerns about overfitting. Further investigation revealed that features like rehabilitation days and grafting were post-admission features, meaning they reflected treatment decisions already made rather than helping to predict initial admission to ICU or the General Ward. Outcome:

We realized that post-admission features were dominating the model, so we moved to exclude those features and refocus the prediction on pre-admission indicators. Attempting Other Models: Refining the Feature Set To improve the model and avoid over-reliance on post-admission indicators, we shifted to a more refined feature set that only included variables relevant to the patient's condition upon admission.

Approach 3: Refined Feature Selection and ICU Prediction

Refined Features:

We excluded features like rehabilitation days and grafting, focusing on: Age (age) Burn Severity (tbsabrn) Ventilator Days (daysvent) Interaction Terms such as: Age × Burn Severity (tbsabrn_age_interaction) Challenges:

After refining the features, the model accuracy dropped to a more reasonable level, but challenges arose in correctly distinguishing between ICU and General Ward patients, particularly for borderline cases. Outcome:

This approach gave us a clearer understanding of the core predictors for ICU admission, but we needed further fine-tuning. Exploring Other Approaches: Survival Rate Prediction We also briefly explored the possibility of predicting survival rate and length of stay based on patient data, but due to time constraints and data limitations, this approach was set aside in favor of improving the ICU vs. General Ward classification.

Final Approach: Predicting ICU vs. General Ward with an 86% Accuracy

Through iterative refinement, we arrived at a model that could predict whether a patient required ICU care with 86% accuracy. Here's a summary of what worked:

Key Features: The most impactful features in determining ICU vs. General Ward were: Ventilator Days (daysvent): Patients requiring ventilation typically needed ICU care. Total Burn Surface Area (tbsabrn): Larger burn areas often correlated with ICU stays. Age (age): Older patients tended to require more intensive care. Model: The final model was a Random Forest Classifier with the following key characteristics: Accuracy: 86% Precision and Recall: These were balanced, with the model correctly identifying most ICU patients while minimizing false positives and false negatives for General Ward patients. Challenges Overcome: Feature Selection: By excluding post-admission features, we ensured the model was based on patient data available at the time of admission, making the predictions more actionable. Overfitting: The initial model suffered from overfitting due to features that reflected decisions already made (e.g., rehabilitation), but refining the feature set resolved this. Key Lessons and Takeaways Over-reliance on Post-admission Features: Features like grafting and rehabilitation days should be excluded from initial admission predictions since they reflect treatment rather than pre-admission conditions.

Feature Importance: The most important factors in predicting ICU care were ventilator days, burn severity, and age. These variables directly relate to a patient's condition upon admission, ensuring a more accurate prediction model.

Balancing Accuracy and Simplicity: We found that focusing on simpler, more intuitive features resulted in an accuracy of 86%, which is both reasonable and actionable in a clinical setting.

Conclusion Our journey began with the aim of predicting hospitalization, but through several iterations and refinements, we landed on a well-balanced model for predicting whether a patient would require ICU or General Ward care. With 86% accuracy, the final model successfully leverages key pre-admission variables like ventilator days, burn severity, and age, ensuring a robust and clinically relevant prediction tool.