Predicting Patient Readmission

Predicting readmission matters because it helps hospitals identify which patients may need stronger discharge planning, earlier follow-up, or additional care coordination. Readmissions are costly, strain hospital capacity, and can reflect gaps in continuity of care. A useful prediction model can support earlier intervention and help direct limited resources toward the patients at highest risk.

Our approach focused on structured encounter and patient information. We created readmission labels by checking whether a patient returned within a fixed time horizon after an encounter, and repeated that setup across 5-, 10-, 15-, 20-, 25-, 30-, 60-, and 90-day windows. Categorical variables such as visit type, admission source, diagnosis group, and demographic fields were target-encoded using out-of-fold encoding, while numeric features were standardized. We then trained three models on the same processed dataset: XGBoost, Random Forest, and a small neural network.

XGBoost performed best at almost horizon. At 5 days, it achieved 0.7062 accuracy and 0.7312 AUC, slightly ahead of Random Forest and the neural model. By 30 days, XGBoost reached 0.7244 accuracy and 0.7428 AUC, and at 90 days it reached 0.8475 accuracy and 0.7792 AUC. Across the tree-based models, visit type was the strongest predictor, with diagnosis and patient birth year also contributing heavily. The visit type became weaker as the horizon got larger. Overall, the results show that readmission risk can be predicted effectively from structured clinical and encounter data, and that boosted trees are the strongest of the methods we tested.

Relationship between Patient Relationship and Time to Followup

This relationship matters because it can indicate the importance of rural hospitals and the option of telehealth in rural vs. urban areas.