Predicting 120-day hospital readmission using medical administrative patient data*

Marijn van Wingerden[1][0000−0002−0558−6624], Jelle de Boer[2,3], and Eric Postma[1]

[1] Cognitive Science & Artifical Intelligence, Tilburg University, Warandelaan 2 Tilburg, the Netherlands e.j.m.vanWingerden@uvt.nl
[2] Master track: Data Science: Business & Governance, Tilburg University, Warandelaan 2 Tilburg, the Netherlands
[3] LOGEX Healthcare Analytics, WTC Amsterdam B-tower, 13th floor Strawinskylaan 1329, Amsterdam

Abstract. Hospitals and health-care insurers routinely use models to predict patient readmission, extrapolating from historical data. Subsequently, the predicted quantities can be used for contracting and pricing negotiations between these hospitals and healthcare insurers. The Dutch healthcare system uses unique standardized Care Trajectories (so-called DBCs) for administration and billing of care. Here, we compared supervised machine learning methods on predicting 120-day readmission as an operationally significant metric. We used administrative patient data from 21 common Care Trajectories, in combination with demographic information. A lightGBM model using undersampling to tackle class imbalance yielded an AUROC score of 0.86 and provided the highest recall score (73.8%).

Keywords: Health Analytics · Hospital Readmission · Care Trajectory · Administrative Patient Data

1 DATASET:

Administrative patient data was obtained from a company that aggregates and analyses patient data for healthcare institutions to provide cost estimates. The data originates from one hospital, was registered between 2016 and 2018 and includes patient demographics, DBC billing codes, billing amounts, and potential readmissions. This set of raw data consisted of approximately 3 million rows, from which 120-day readmission as a target label was extracted. After pre-processing to include only DBCs with more than 1000 cases, the remaining 790630 rows consisted of 273,345 (34.6%) readmitted and 517,285 (65.4%) not readmitted patients.

2 RESEARCH QUESTION:

Can 120-day hospital readmission be meaningfully predicted from patient demographics and care trajectory billing information? Which type of machine learning method performs best in this prediction?

3 METHODS:

Feature selection was performed on a separate subset of the data by ranking on importance using Gradient Boosting Trees and Random Forest with Gini-Impurity/Information Gain. After ranking the features in decreasing order, the five features with least importance were removed in an iterative procedure. By evaluating the average AUROC-score, we found that the optimal number of features was 30. These features were used to train Logistic Regression, Decision Tree, Random Forest, XGBoost, LightGBM models, and through an ANN with four (hidden) dense layers. Performance was assessed for each classifier using 5-fold cross validation. To compare hospital-wide models with diagnosis/care trajectory-specific models, we selected 180 DBCs containing more than 1000 cases (see Fig. 1) and split the data (step 1) to train/evaluate (step 2) models both on the whole dataset (step 3.1) and on just the DBC-specific data (step 3.2), and compared the performance for these 180 DBCs.

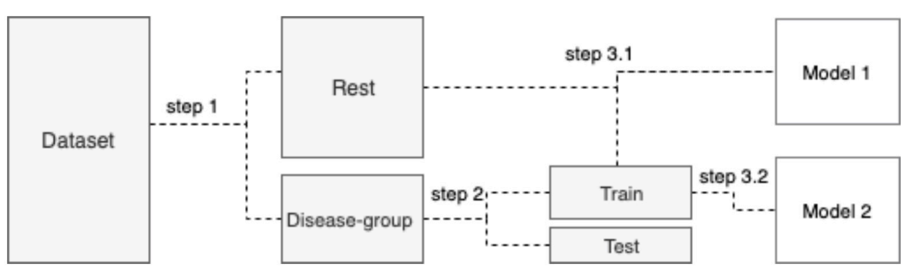

Fig. 1. Preparing training and test data for a hospital-wide and disease-specific model. Data is split to consider only 1 DBC (disease group) or the whole dataset (Step 1). Models are trained/evaluated (Step 2) separately for the hospital wide (Step 3.1) and DBC-specific (Step 3.2) models.

4 RESULTS:

We found that adding features, ranked on importance, improved AUROC scores up to around 30 features. The most important features were whether a patient was in their first care trajectory, or a follow-up care trajectory, and the number of days since the last care activity in the care trajectory, suggesting that these DBC-related data points can significantly improve prediction performance above using only demographic information. We found that the lightGBM outperformed all other ML classifiers, arriving at a recall score of 73.8% after implementing undersampling to combat class imbalance. Importantly, locally (disease /caretrajectory-specific) trained models outperformed hospital-wide models in almost all cases. ANNs performed best on homogeneous DBC groups with F1-scores ranging from 0.53 to 0.72 over 6 groups of DBCs, suggesting an advantage in using the ANN over ML-methods when training on more homogeneous groups of patients (see Table 1). Importantly, models trained on DBCs with high re-admission rates (more than 50%) performed much better (median F1: 0.82) compared to models trained on DBCs with low re-admission rates (less than 50%; median F1: 0.55).

Table 1. F1-scores: Diagnosis-specific vs. Hospital-wide models; ANN vs. LightGBM. RR: Readmission Ratio; NTest: Number of observations in test set

Diagnosis	ANN		LightGBM		NTest	RR
	Local	Wide	Local	Wide		
COPD	**0.6822**	0.6611	0.6216	0.5893	2355	0.46
Heart Failure	**0.6913**	0.6685	0.6382	0.6261	1578	0.44
Pneumonia	**0.5393**	0.5119	0.4195	0.4552	569	0.28
Acute Myocardial Infarction	**0.5477**	0.5096	0.2425	0.5272	308	0.34
Osteoarthritis Knee	0.7276	**0.7295**	0.7214	0.6668	1461	0.47
Osteoarthritis Hip	**0.6607**	0.6409	0.6124	0.5903	2208	0.41

5 DISCUSSION:

As the number of (un)planned hospital readmissions is used as a hospital-wide indicator for both quality of care and financial outcomes, obtaining accurate predictions for this metric is essential. We found that a LightGBM – an ensemble method, outperformed the other machine learning methods considered. Promising results were found for a diagnosis-specific trained ANN. Feature ranking and selection suggested that DBC care trajectory-specific information contributed significantly to model performance. From a cost projection perspective, we found that models performed better for care trajectories with higher readmission rates (associated with higher costs) and thus could contribute to more accurate cost predictions.