Machine Learning models for predicting 30-day readmission of elderly patients using custom target encoding approach

N. Nazyrova, T. J. Chaussalet, S. Chahed

School of Computer Science and Engineering, University of Westminster, London W1W 6UW, UK
W1169140@my.westminster.ac.uk

Abstract. The readmission rate is an important indicator of the hospital quality of care. With the upsetting increase in readmission rates worldwide, especially in geriatric patients, predicting unplanned readmissions becomes a very important task, that can help to improve the patient's well-being and reduce healthcare costs. With the aim of reducing hospital readmission, more attention is to be paid to home healthcare services, since home healthcare patients on average have more compromised health conditions. Machine Learning and Artificial intelligence algorithms were used to develop predictive models using MIMIC-IV repository. Developed predictive models account for various patient details, including demographical, administrative, disease-related and prescription-related data. Categorical features were encoded with a novel customized target encoding approach to improve the model performance avoiding data leakage and overfitting. This new risk-score based target encoding approach demonstrated similar performance to existing target encoding and Bayesian encoding approaches, with reduced data leakage, when assessed using Gini-importance. Developed models demonstrated good discriminative performance, AUC 0.75, TPR 0.69 TNR 0.67 for the best model. These encouraging results, as well as an effective feature engineering approach, can be used in further studies to develop more reliable 30-day readmission predictive models.

Keywords: 30-day readmission, home care patient readmission, machine learning model, categorical feature encoding, customized target encoding

1 Introduction

Readmission to the hospital within 30 days from discharge has been receiving growing attention due to its implications on cost and quality of care. In the UK, approximately one in six hospital admissions result in readmission, and elderly patients are more likely to be at risk of readmission [1]. Patients aged 65 and older account for 56% of readmission cases, which constitutes 60% of associated costs [2]. Moreover, there is a growing number of patients with multi-morbidities and in the future the comorbidities seem to take on greater importance due to the overall world's population ageing [3]. In the USA,

ICCS Camera Ready Version 2022
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-08757-8_12

30-day readmissions were higher among elderly Medicare beneficiaries with chronic conditions (22.5%) than among those with acute conditions (19.3%) [4].

But not all readmission cases are unavoidable: the Medicare Payment Advisory Commission (MedPac) estimates 12% of readmission as potentially avoidable and prevention of 10% of these cases could save Medicare $1 billion [5]. Simple post-discharge calls and follow-up visits have proven to be effective measures to decrease early readmission cases in elderly patients [6]. With the aim of reducing ill health and preventing emergency admissions more attention should be paid to care home patients. Older people living in care homes are among the highest risk group for preventable ill health and the use of clinical services [7].

In the past decade, various efforts were invested in modelling the risk of 30-day readmission to hospitals. Several risk scoring systems are widely developed and adopted in hospitals to predict the risk of readmission or mortality. These scoring systems are based on baseline information obtained during the patient's hospital stay. But there is a lack of predictive models that consider the impact of prescribed medications along with more detailed clinical data.

In this paper we describe the development of the hospital readmission model which is based on the thorough health status of the patient created for a full cohort of elderly patients and a subset of home-care patients. In the feature engineering step, we developed a novel score-based target encoding approach. Our categorical feature encoding method increased the statistical performance of the model when compared to target agnostic encoding approaches. The models are built on the clinical data which is available before discharge, hence can be used to predict the risk of patient readmission and undertake preventive measures.

2 Study Design and Methodology

The study was a cross-sectional assessment of 63 557 geriatric patients which constitutes 140 518 hospital admissions between 2015 and 2019 from the Medical Information Mart for Intensive Care (MIMIC-IV) datasets to predict 30-day readmission.

We defined the criteria for readmission as an episode of unplanned hospitalization to an acute care hospital within 30 days of previous discharge. The unplanned hospitalizations are recorded under the emergency and urgent admission types, which can include both walk-in admissions and emergency department admissions. Whereas the planned hospitalizations are recorded as observation or elective admission types. Only unplanned hospitalizations were included in the analysis. Moreover, to exclude the episodes of observation stays in the emergency department, patients who spent less than one day in the emergency department are not considered readmitted.

Fig. 1 shows the methodology adopted in this work. In the data preparation step, the clinical dataset for the study was extracted. The analysis was performed on the MIMIC-IV dataset - a large database with administrative, clinical and critical care data for patients admitted to the Beth Israel Deaconess Medical Centre [8]. It contains data about over 382 278 deidentified patients, which constitutes 523 740 hospital admissions. All

ICCS Camera Ready Version 2022
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-08757-8_12

patient identifiers are removed to comply with the Health Insurance Portability and Accountability (HIPAA) regulations [9].

In the data pre-processing step, the extracted dataset was split into training and test sets with 70% and 30% of records correspondingly. The missing value imputation and standardisation were applied to each set separately to avoid data leakage. The training set was modified with a state-of-the-art oversampling technique and both imbalanced and balanced datasets were used for the modelling using seven ML algorithms. In the feature engineering step, various categorical features were compared, and the novel score-based categorical feature encoder was proposed. Mean Decrease Impurity Filter was used for Feature Selection and feature importance levels were monitored with Gini Purity Scores. For the model evaluation, the AUC of the models and classes' accuracy rates were compared after the testing phase.

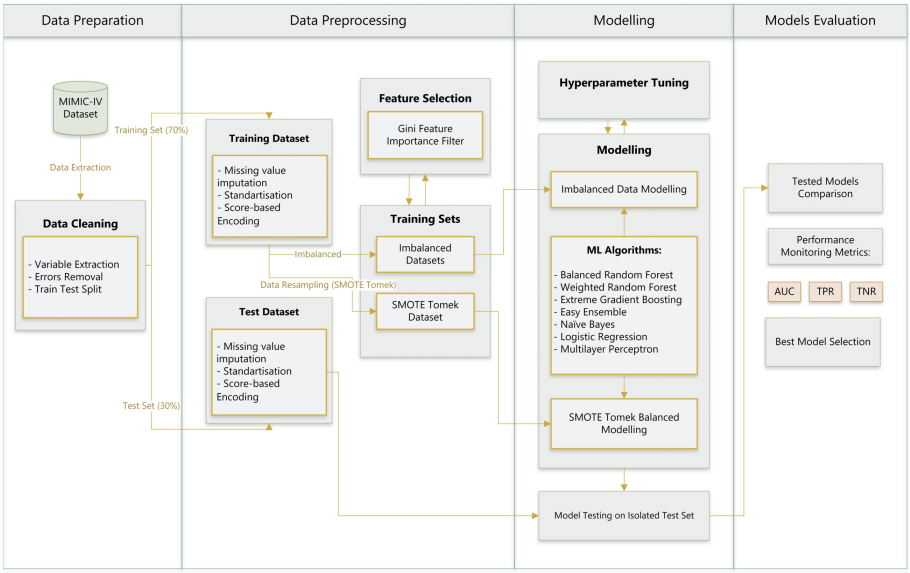

Fig. 1. The ML modelling methodology adopted in this work

2.1 Data preparation

The full MIMIC-IV cohort had 523 740 admissions and 382 278 patients, which were further filtered out based on several criteria. For the analysis, only patients older than 65 years were included. Moreover, since the readmission rate is used to evaluate the quality of care in the hospitals, those patients who were discharged against advice and those who were discharged to hospices were excluded. The full sampling approach is demonstrated in fig. 2. For the study, the final cohort of 140 518 admissions was used, of which 18 447 (13.12%) admissions were identified as readmission cases.

ICCS Camera Ready Version 2022
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-08757-8_12

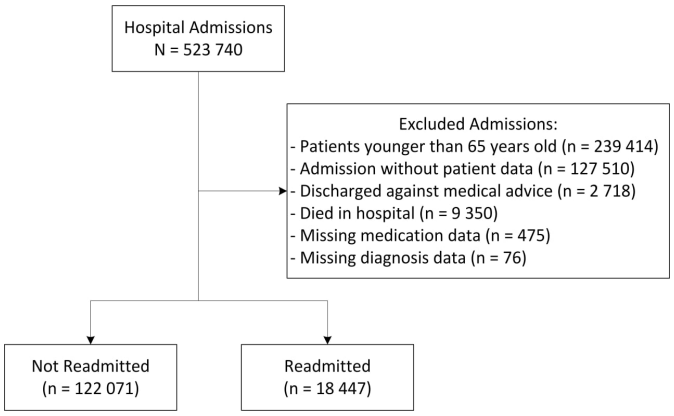

Fig. 2. Study flow diagram

2.2 Data pre-processing

To prepare the datasets for the modelling and enhance performance several data transformations were performed.

Missing value imputation
Demographical variables often had a high number of missing values, therefore additional dimensions to indicate unknown features were added. A small number of patients (0.6%) did not have any associated diagnoses and medications records, therefore were excluded from the analysis. In 5% of admissions laboratory values for haemoglobin and sodium were missing. These variables were considered within normal limits, following the haemoglobin and sodium value imputation approach in similar studies [10][11]. Most of the features in the newly formed dataset are dichotomous, hence only zero imputation was made for missing values.

Feature Scaling
All input variables were scaled to a common magnitude using a robust scaling approach [12]. A robust scaler removes the median and scales the data between the 25^{th} quantile and 75^{th} quantile range. Since the medical records often contain outliers, especially for geriatric patients with multiple comorbidities, the mean can be skewed by the extreme values, and typically these extreme values have a low probability of occurrence. Therefore, a robust scaling approach was adopted to avoid the negative impact of outliers on standardization.

Feature engineering
Most machine learning algorithms required the input data to be a numeric matrix, hence it is required to encode categorical features, that do not have an intrinsic ordering. There are numerous ways to encode categorical features, however, not all of them preserve the original knowledge contained in categorical features. To select the most suitable categorical feature encoder, target encoding, weight of evidence encoder and custom-made score-based target encoder are compared.

ICCS Camera Ready Version 2022
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-08757-8_12

Target encoding
With target encoding, each category is replaced with the mean target value for samples having that category [13]. The target value is the y-variable or the value the model is trying to predict [13]. This allows encoding categories without increasing the data dimensionality preserving the original information of the features. This approach is performing particularly well on large amounts of training data and categorical features with low cardinality. For each category the average value of the target label is calculated on the training examples. Further, the mean encoding is mapped to the test set. However, this approach is often criticized for the tendency to overfit due to the target leakage [13]. In addition, when categories have few training examples, mean target values for these categories may be not representative, deteriorating the model performance.

Weight of evidence encoder.
Weight of Evidence is a categorical feature encoder that measures the strength of a grouping technique that is used to separate one class from another in the following way: [14].

$$WoE = ln\left(\frac{\%\ of\ non{-}events}{\%\ of\ events}\right) \qquad (1)$$

Similarly to Target Encoding, there is a potential for target leakage and overfitting of the model. To avoid this, random Gaussian noise may be injected to the variable during encoding.

Score-based Target Encoding
Since the target encoding and weight of evidence encoding are often criticized for data leakage and overfitting, it was decided to adjust the encoder to avoid this behaviour. Similar to target encoding, categorical features are encoded using the target value in the training set. Features are replaced with the blend of the posterior probability of the target given a particular categorical value and the prior probability of the target over all the training data.

The readmission rate (target variable ratio) is used as a baseline rate, which is 13.12% for our dataset.

$$Baseline\ Rate = \frac{Readmitted}{Readmitted + Not\ Readmitted} * 100 \qquad (2)$$

When the feature readmission rate is within ½ standard deviation from the baseline readmission rate, this feature is encoded to 1, following the risk-scoring approach. Score boundaries were calculated using *m* standard deviations from the baseline readmission rate, where *m* is the incremental value.

```
If Encoded(X^i) == Baseline Rate ± 0.5 σ
  Then Score(X^i) = 1
If Encoded(X^i) == Baseline Rate + (m+0.5)σ
  Then Score (X^i) feature = m+1
If Encoded(X^i) == Baseline Rate – (m+0.5)σ
  Then Score (X^i) feature = 1-m
```

ICCS Camera Ready Version 2022
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-08757-8_12

Where: $(\mathcal{X}^i)$ is given categorical feature, Encoded $(\mathcal{X}^i)$ is the probability of the target (readmitted) given particular categorical value $(\mathcal{X}^i)$ and the prior probability of the target over all the training data, σ - standard deviation of Encoded$(\mathcal{X}^i)$ for the given categorical feature $(\mathcal{X}^i)$, m – number of standard deviations;

To demonstrate how custom score-based target encoding is used, the example of encoding for 'Discharge Location' categorical value is provided in Table 2.

Table 1. Example of score-based encoding using 'Discharge Location' categorical feature

Discharge Location	Not Readmitted	Readmitted	Readmission % for the feature	Encoded Value
psych facility	209	134	39.06	9
chronic/long term acute care	3745	1039	21.71	4
other facility	161	31	16.14	2
acute hospital	631	109	14.72	1
home health care	37620	6369	14.47	1
skilled nursing facility	31409	5117	14.00	1
assisted living	430	66	13.30	1
rehab	5881	884	13.06	1
home	41939	4687	10.05	0
Standard Deviation			**3.339**	

This approach can be strongly affected by the outliers in the standard deviation calculation. Therefore, the outliers should be omitted when calculating standard deviation. To avoid data leakage encoding should be based on the training set data and the obtained scores should be mapped to categorical features in the test set.

There was no significant difference in the statistical performance of the three encoders during the testing as can be seen in Table 1. However, since the problem of data leakage in target encoding and weight of evidence encoding was raised, we decided to observe the feature importance of these encoded categorical features (Fig. 3).

ICCS Camera Ready Version 2022
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-08757-8_12

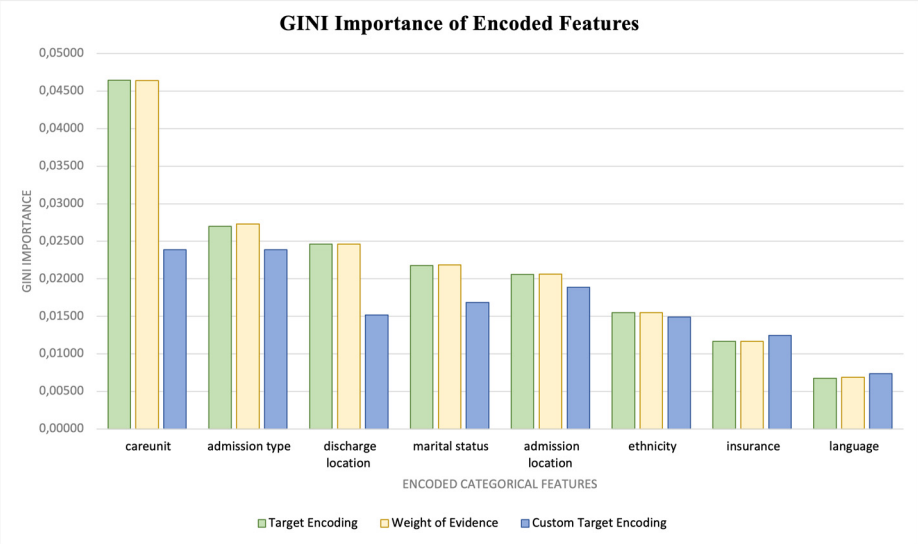

Fig. 3. Gini Importance of encoded features.

Both target encoding and weight of evidence encoding demonstrate a higher variance in feature importance than custom target encoding. Particularly, the care unit categorical feature which contained a large number (31) of categories, has the highest feature importance of 0.46 for both target and weight of evidence encoders. This can be an indicator of data leakage for high cardinality features. Surprisingly, the majority of low cardinality features also have higher feature importance in target and weight of evidence encoders. The custom target encoding approach demonstrates a low variance of feature importance, at the same time models with this encoder demonstrate slightly higher AUC (0.06 increase) compared to the next best model built with the target encoder. The score-based target encoder makes the training algorithms put less emphasis on the categorical features, thus reducing target leakage. Table 1. demonstrates the performance of tree-based and linear models with the target, weight of evidence and score-based target encoding.

Table 2. Performance Metrics of Machine Learning Models with three different categorical feature encoding approaches: target, weight of evidence and score-based target encoding.

	Target Encoding			**Weight of Evidence Encoding**			**Score based target encoding**		
Models	**AUC**	**TPR**	**TNR**	**AUC**	**TPR**	**TNR**	**AUC**	**TPR**	**TNR**
Balanced Random Forest	0.7414	0.68	0.67	0.7429	0.68	0.67	0.7409	0.68	0.67
XGBoost	0.7485	0.69	0.66	0.7493	0.69	0.67	0.7501	0.69	0.67
Logistic Regression	0.7213	0.64	0.69	0.7302	0.65	0.68	0.7212	0.64	0.69
Naïve Bayes	0.6913	0.62	0.66	0.6914	0.62	0.66	0.6913	0.62	0.66

ICCS Camera Ready Version 2022
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-08757-8_12

Data resampling

To tackle the class imbalance Hybrid SMOTE-Tomek Links resampling approach is used. The minority class is oversampled using SMOTE algorithm: creating artificial instances based on k-nearest neighbours. When the dataset contains ambiguous records, specifically, with two closest neighbour instances belonging to the opposite classes, such records are removed using the Tomek Links approach [15]. This approach helps to improve the class separation near the decision boundaries.

Feature selection

Mean Decrease in Impurity was used to monitor the feature importance and select features for the model. Features with zero importance were excluded from the analysis. Fig. 4. shows the top features with their Gini importance that were selected for the analysis.

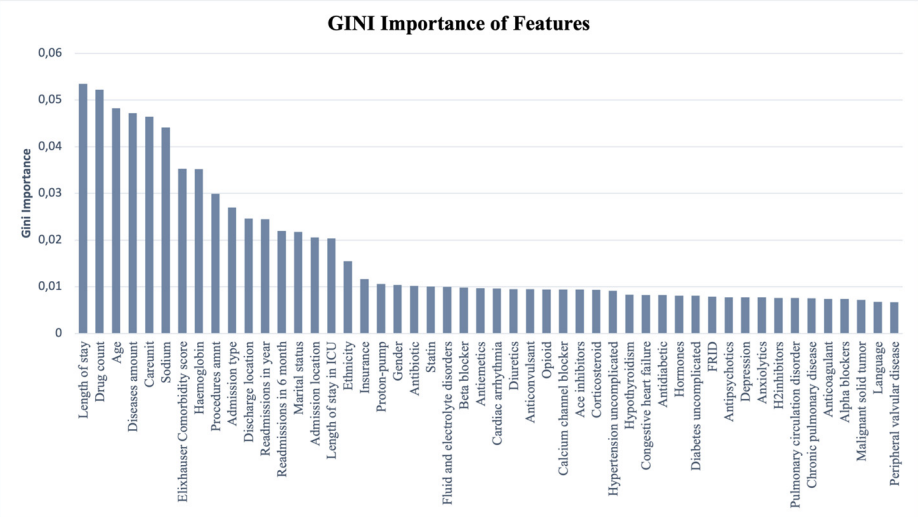

Fig. 4. Feature importance based on Gini Impurity

Since readmission modelling implies a higher level of uncertainty, more complete information about the patients should be used to improve the model discrimination. Several studies [16][17] suggested that polypharmacy and medication noncompliance are one of the major reasons for readmission, followed by fall injuries when older patients are concerned. Therefore, it was decided to incorporate into the analysis these cases. Thirty-two groups of drugs that have the highest correlation to readmission were identified based on the existing studies [16][17]. These drug groups included mainly drugs that increase the risk of falls, such as antihypertensives, antiarrhythmics, anticholinergics, antihistamines, sedatives, antipsychotics, and opioids. Moreover, some studies [18] outlined the increased risk of readmission for patients taking heart failure drugs,

ICCS Camera Ready Version 2022
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-08757-8_12

such as ace inhibitors, angiotensin receptor blockers, beta-blockers, and diuretics. The final model contains twenty-two groups of drugs chosen during the feature selection.

Even though most of the existing readmission studies use the Charlson Comorbidity index features, for this study, it was decided to use the Elixhauser Comorbidity index features, as it gives a more detailed view of patients' health conditions [19]. Elixhauser Comorbidity score incorporates 31 groups of diseases, including those in the Charlson comorbidity index. The full list of attributes used for modelling is provided in Table 3.

Table 3. Model Variables

Attributes Groups	Attributes
Demographic variables	Age, gender, ethnicity, marital status, language
Administrative variables	Length of stay, admission type, insurance, number of readmissions during the previous year, number of readmissions during the previous six month, number of ICD9/10-coded procedures, number of ICD9/10 encoded diseases, number of prescribed drugs, discharge care unit
Elixhauser comorbidity score diseases	Congestive heart failure, cardiac arrhythmias, valvular disease, pulmonary circulation disorders, peripheral vascular disorders, hypertension, paralysis, neurodegenerative disorders, chronic pulmonary disease, diabetes uncomplicated, diabetes complicated, hypothyroidism, renal failure, liver disease, peptic ulcer disease without bleeding, AIDS/HIV, lymphoma, metastatic cancer, solid tumour without metastasis, rheumatoid arthritis, coagulopathy, obesity, weight loss, fluid and electrolyte disorders, blood loss anaemia, deficiency anaemia, alcohol abuse, drug abuse, psychosis, depression
Prescribed medications (25 drug groups)	Anticoagulants, antibiotics, antipsychotics, anticonvulsants, anticholinergics, antiarrhythmics, antiemetics, anti-diabetic, antifungal, angiotensin, alpha-blockers, anxiolytics, ACE inhibitors, beta-blockers, calcium channel blockers, cardiac drugs, corticosteroids, diuretics, FRIDs (fall risk increasing drugs), h2 inhibitors, hormones, mineralocorticoids, PIMs (potentially inappropriate drugs), sedatives, h2 inhibitors.

2.3 Modelling

Six machine learning and one deep learning algorithms were used to build predictive models: Logistic Regression, Naïve Bayes, Random Forest (balanced and weighted), Easy Ensemble, Extreme Gradient Boosting and Multilayer Perceptron. To account for the class imbalance, algorithms hyper-parameters to adjust class weights were applied. Models were validated on the training set using 5-fold validation. Models were tested on imbalanced dataset settings and Smote Tomek balanced dataset settings.

2.4 Model Evaluation.

The selection of appropriate evaluation metrics is very important in clinical decision making and it often implies a trade-off between cost optimization and risk aversion. Acknowledging these risks, it was decided to use the following evaluation metrics: the

ICCS Camera Ready Version 2022
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-08757-8_12

True Positive Rate (Sensitivity), the True Negative Rate (Sensitivity) and the Area Under Receiver Operating Characteristic Curve (AUROC).

True Positive Rate (TPR) is used to measure the rate of correct prediction of the subjects from the readmission group. And **True Negative Rate (TNR)** is used to measure the rate of correct prediction of the subjects from the non-readmission group. ROC AUC metrics is a commonly used evaluation metric is a graph showing the performance of a classification model at all classification thresholds. This metric can show a trade-off between correctly predicted readmission cases and those who were falsely classified as readmitted, which in turn can impose additional costs on healthcare organizations. To select the best models the highest AUC score was selected with the most balanced TPR and TNR. We assume equal classification importance for both classes.

3 *RESULTS*

Imbalanced Classification results

Table 4 shows the test performance of models built with the imbalanced dataset settings for all elderly patients and for those who were discharged to home health care settings. Multilayer Perceptron achieved the lowest AUC score of 0.61, and 0.26 TPR, 0.85 TNR due to the class imbalance. There are no hyper-parameters to account for the class imbalance. Ease Ensemble learner achieved 0.73 AUC, however the class accuracy difference achieved as much as 10%. Traditional ML algorithms showed less variance between TPR and TNR when used in a cost-sensitive approach with balanced class weights. The difference between the class accuracy was less than 5% for the remaining models. Naïve Bayes achieved the second lowest result with AUC 0.69, followed by the Logistic Regression. Tree-based ensemble learner demonstrated the best statistical performance, with the XGBoost achieving the highest 0.75 AUC with the most balanced class accuracy: 0.69 TPR and 0.67 TNR for the dataset with a full cohort of patients, and AUC 0.74, TPR 0.68 and TNR 0.65 for the home care patients.

Table 4. The test performance of ML/AI models trained with imbalanced datasets

ML Models	**Full cohort of patients**			**Home Care patients**		
	AUC	TPR	TNR	AUC	TPR	TNR
Balanced Random Forest	0.74	0.68	0.67	0.73	0.67	0.65
XGBoost	**0.75**	**0.69**	**0.67**	**0.74**	**0.68**	**0.65**
Weighted Random Forest	0.72	0.64	0.64	0.71	0.64	0.68
Easy Ensemble	0.74	0.67	0.57	0.73	0.65	0.57
Logistic Regression	0.72	0.64	0.69	0.71	0.68	0.62
Naïve Bayes	0.69	0.62	0.66	0.68	0.59	0.66
Multilayer Perceptron	0.61	0.26	0.85	0.61	0.25	0.85

SMOTE Re-sampled Datasets Modelling Results

ICCS Camera Ready Version 2022
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-08757-8_12

SMOTE Tomek resampling technique applied to the imbalanced dataset did not result in improved model performance. Overall, there is a slight AUC score decrease for most of the models with the resampled dataset settings. The class variance was reduced as expected. Multilayer Perceptron has less class accuracy imbalance, and still demonstrates the poorest AUC and TPR. Other predictors demonstrate a balanced classification between the two classes. XGBoost again demonstrates the highest AUC and TPR for both dataset settings: full cohort of patients and home care patients (see Table 5).

Table 5. The test performance of ML/AI models trained with resampled datasets

ML Models	**Full cohort of patients**			**Home Care patients**		
	AUC	TPR	TNR	AUC	TPR	TNR
Balanced Random Forest	0.73	0.60	0.67	0.72	0.65	0.65
XGBoost	**0.74**	**0.71**	**0.64**	**0.73**	**0.69**	**0.62**
Weighted Random Forest	0.72	0.64	0.64	0.71	0.64	0.68
Easy Ensemble	0.71	0.66	0.57	0.70	0.65	0.59
Logistic Regression	0.71	0.63	0.68	0.70	0.66	0.62
Naïve Bayes	0.69	0.63	0.66	0.69	0.67	0.64
Multilayer Perceptron	0.65	0.30	0.79	0.64	0.30	0.78

Baseline characteristics analysis

Analysis of baseline characteristics of the baseline population demonstrated that overall home care patients have more aggravated health conditions, with the average age higher than the full cohorts', they have more comorbidities and have a longer length of stay in the hospital. They are prescribed more medications and have higher number of registered diseases. Moreover, this cohort of patients more often live without partner (widowed or divorced). There are more women admitted to the hospital, however, men more often have readmission cases. Detailed baseline characteristics are provided in the Table 6.

Table 6. Baseline characteristics of the study population by 30-day readmission status.

Characteristics	**Full Cohort**		**Home Care patients**	
Readmission (%)	Not Readmitted 122,071 (87%)	Readmitted 18,447 (13%)	Not Readmitted 69,042 (86%)	Readmitted 11,490 (14%)
Age, (mean, SD)	77.06 (8.23)	76.89 (8.18)	78.70 (8.42)	78.14 (8.36)
Length of Stay (mean, SD)	5.13 (6.28)	6.43 (8.3)	5.80 (6.02)	6.46 (7.73)
Emergency department visits in last 12 month (mean, SD)	0.29 (0.82)	0.91 (1.79)	0.34 (0.90)	0.98 (1.88)
Emergency department visits in last 6 month (mean, SD)	0.23 (0.66)	0.73 (1.35)	0.27 (0.73)	0.77 (1.41)
Gender:				
Male	58 752 (48%)	9 398 (51%)	30 868 (45%)	5 388 (48%)
Female	63 319 (52%)	9 049 (49%)	38 429 (55%)	5 865 (52%)
Ethnicity:				
White	92 743 (75%)	13 547 (76%)	51 930 (75%)	8 644 (75%)
Asian	3 451 (3%)	527 (3%)	1 792 (2%)	323 (2%)

ICCS Camera Ready Version 2022
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-08757-8_12

Hispanic/Latino	3 811 (3%)	606 (3%)	2 089 (3%)	421 (3%)
Black/African American	14 065 (11%)	2 293 (13%)	8 245 (12%)	1 620 (14%)
American Indian/Alaska Native	214 (0.2%)	37 (0.2%)	117 (0.16%)	26 (0.23%)
Unknown	8931 (7%)	768 (4%)	4869(7%)	456 (4%)
Insurance:				
Medicaid	2,241 (2%)	352 (2%)	938 (1.5%)	186 (2%)
Medicare	83 291 (67.5%)	12 374 (70%)	47 942 (69.5%)	8 115 (71%)
Other	37 683 (30%)	5 052 (28%)	20 162(29%)	3 119 (27%)
Marital Status:				
Married	60 022 (49%)	8 678 (49%)	29 832 (43%)	5 074 (44%)
Single	23 021 (18%)	3374 (19%	13 390 (19%)	2 268 (20%)
Widowed	27 991 (22%)	4 368 (24%)	18 712 (27%)	3 258 (28%)
Divorced	9 139 (7%)	1 234 (7%)	5,323 (8%)	824 (7%)
Unknown	3 042 (2%)	124 (0.7%)	1,785 (2%)	66 (0.5%)
Language:				
English speaking	107 584 (87.4%)	15 342 (86%)	59 814 (87%)	9 772 (85%)
Non-English Speaking	15 631 (12.6%)	2 436 (14%)	9 228 (13%)	1 718 (15%)
Admission type:				
Emergency	65 628 (54%)	13 147 (73%)	36 324 (53%)	8 265 (72%)
Urgent	11 651 (9%)	1 944 (11%)	7 058 (10%)	1 166 (10%)
Elective	44 792 (37%)	3 356 (17%)	25 650 (37%)	2 059 (18%)
Charlson Comorbidity Index (mean, SD)	3.19 (2.99)	4.10 (3.13)	3.45 (3.08)	4.36 (3.16)
Diseases Amount (mean, SD)	14.44 (6.97)	15.96 (7.20)	15.67 (6.91)	16.72 (7.00)
Prescribed drugs amount (mean, SD)	23.11 (11.22)	26.24 (12.41)	25.11 (10.87)	26.61 (10.90)

4 Discussion

We developed machine learning and artificial intelligence models to predict the 30-day hospital readmission. The developed model was validated on the full cohort of elderly patients and a subset of patients, who use home care services.

We adopted a novel customized target encoding approach in the feature engineering step. Categorical variables were encoded using a score-based target encoder. This approach demonstrated similar performance with the target encoder and the weight of evidence encoder. However, a comparison of Gini feature importance for all three types of encoders upholds the well-known problem of data leakage in target encoders, including weight of evidence encoder. Whereas score-based target encoding demonstrated less emphasis on encoded categorical features, preserving the good model performance. Developed models were modified to account for the high-class imbalance. While traditional algorithms benefited from the hybrid oversampling approach, tree-based models performed better in ensemble learner mode with the class weight adjustment.

Among all developed and analysed models, ensemble learners and specifically XGB dominated the highest ranking in TPR, TNR and AUC scores for both imbalanced and re-sampled data states for the proposed dataset with AUC 0.75, TPR 0.69, TNR 0.67.

ICCS Camera Ready Version 2022
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-08757-8_12

During our analysis we found that polypharmacy is one of the most important predictors of readmission. A higher number of prescribed medications had a positive correlation with readmission cases. Moreover, some drug groups were identified to have an impact on classification results, such as diuretics, corticosteroids, anticoagulants and fall risk-increasing drugs.

In concordance with previous studies, the number of comorbidities directly correlates with the frequency of 30-day readmission after hospital discharge [20][21]. Particularly such conditions as fluid and electrolyte disorders, cardiac arrhythmia, hypertension, congestive heart failure, pulmonary circulation disorders, depression, and malignant cancer. Interestingly, some of these comorbidities are not covered in the widely used Charlson Comorbidity index, such as fluid and electrolyte disorder, cardiac arrhythmia, hypertension and depression. Whereas most of the existing studies utilize Charlson Comorbidity Index, Elixhauser Comorbidity Index could provide a better overview of the patients' health conditions [19]. Future work will take into consideration drug-drug interactions as an important readmission predictor. Polypharmacy and drug interactions should be thoroughly examined when analysing geriatric patient readmission. Moreover, future work should consider the impact of post-discharge services on the 30-day readmission when the data is available.

References

1. NHS Digital. Hospital Admitted Patient Care Activity 2019-20, https://digital.nhs.uk/data-and-information/publications/statistical/hospital-admitted-patient-care-activity/2019-20; 2020, last accessed 2022/2/12
2. Glans, M., Kragh Ekstam, A., Jakobsson, U. *et al.:* Risk factors for hospital readmission in older adults within 30 days of discharge – a comparative retrospective study. *BMC Geriatr* 20, 467 (2020). https://doi.org/10.1186/s12877-020-01867-3
3. Fabbri, E., Zoli, M., Gonzalez-Freire, M., Salive, M. E., Studenski, S. A., & Ferrucci, L.: Aging and Multimorbidity: New Tasks, Priorities, and Frontiers for Integrated Gerontological and Clinical Research. *Journal of the American Medical Directors Association*, *16*(8), 640–647 (2015). https://doi.org/10.1016/j.jamda.2015.03.013
4. Raval, A. D., Zhou, S., Wei, W., Bhattacharjee, S., Miao, R., & Sambamoorthi, U.: 30-Day Readmission Among Elderly Medicare Beneficiaries with Type 2 Diabetes. *Population health management*, *18*(4), 256–264 (2015). https://doi.org/10.1089/pop.2014.0116
5. McIlvennan, C. K., Eapen, Z. J., & Allen, L. A.: Hospital readmissions reduction program. *Circulation*, *131*(20), 1796–1803. (2015). https://doi.org/10.1161/CIRCULATIONAHA.114.010270
6. Vernon, D., Brown, J. E., Griffiths, E., Nevill, A. M., & Pinkney, M.: Reducing readmission rates through a discharge follow-up service. *Future healthcare journal*, *6*(2), 114–117. (2019). https://doi.org/10.7861/futurehosp.6-2-114
7. Dixon, J.: Reducing emergency admissions from care homes: a measure of success for the NHS long-term plan. (2019). https://www.hsj.co.uk/emergency-care/reducing-emergency-admissions-from-care-homes-a-measure-of-success-for-the-nhs-long-term-plan/7025675
8. Johnson, A., Bulgarelli, L., Pollard, T., Horng, S., Celi, L. A., & Mark, R.: MIMIC-IV (version 0.4). PhysioNet (2020). https://doi.org/10.13026/a3wn-hq05.

ICCS Camera Ready Version 2022
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-08757-8_12

9. Institute of Medicine (US) Committee on Health Research and the Privacy of Health Information: The HIPAA Privacy Rule; Nass SJ, Levit LA, Gostin LO, editors. Beyond the HIPAA Privacy Rule: Enhancing Privacy, Improving Health Through Research. Washington (DC): National Academies Press (US); 2009. 4, HIPAA, the Privacy Rule, and Its Application to Health Research. Available from: https://www.ncbi.nlm.nih.gov/books/NBK9573/
10. Donzé, J., Aujesky, D., Williams, D., Schnipper, J.L.: Potentially Avoidable 30-Day Hospital Readmissions in Medical Patients: Derivation and Validation of a Prediction Model. *JAMA Intern Med.* 2013;173(8):632–638. doi:10.1001/jamainternmed.2013.3023
11. Robinson, R., Hudali, T.: The HOSPITAL score and LACE index as predictors of 30 day readmission in a retrospective study at a university-affiliated community hospital. PeerJ. 2017;5:e3137. Published 2017 Mar 29. doi:10.7717/peerj.3137
12. Iglewicz, B.: Robust scale estimators and confidence intervals for location. In: Understanding Robust and Exploratory Data Analysis, D. C. Hoaglin, M. Mosteller, and J. W. Tukey, Eds. NY: Wiley (1983).
13. Pargent, F., Pfisterer, F., Thomas, J., & Bischl, B. (2021). Regularized target encoding outperforms traditional methods in supervised machine learning with high cardinality features. arXiv preprint arXiv:2104.00629.
14. Lund, B. (2016). Weight of Evidence Coding and Binning of Predictors in Logistic Regression.
15. Batista, G., Prati, R., Monard, M.: A study of the behavior of several methods for balancing machine learning training data. SIGKDD Explor. 2004;6(1):20–29. doi: 10.1145/1007730.1007735.
16. Glans M, Kragh Ekstam A, Jakobsson U, Bondesson Å, Midlöv P (2021) Medication-related hospital readmissions within 30 days of discharge—A retrospective study of risk factors in older adults. PLoS ONE 16(6): e0253024. https://doi.org/10.1371/journal.pone.0253024
17. Uitvlugt, Elien & Janssen, Marjo & Siegert, Carl & Kneepkens, Eva & Bemt, Bart & van den Bemt, Patricia & Karapinar-Carkit, Fatma. (2021). Medication-Related Hospital Readmissions Within 30 Days of Discharge: Prevalence, Preventability, Type of Medication Errors and Risk Factors. Frontiers in Pharmacology. 12. 567424. 10.3389/fphar.2021.567424.
18. Parajuli P, Lara-Garcia OE, Regmi MR, Skoza W, Bhattarai M, Kulkarni A, Robinson RL. Heart Failure Drug Class Effects on 30-Day Readmission Rates in Patients with Heart Failure with Preserved Ejection Fraction: A Retrospective Single Center Study. Medicines (Basel). 2020 May 20;7(5):30. doi: 10.3390/medicines7050030. PMID: 32443705; PMCID: PMC7281589.
19. Buhr RG, Jackson NJ, Kominski GF, Dubinett SM, Ong MK, Mangione CM. Comorbidity and thirty-day hospital readmission odds in chronic obstructive pulmonary disease: a comparison of the Charlson and Elixhauser comorbidity indices. BMC Health Serv Res. 2019 Oct 15;19(1):701. doi: 10.1186/s12913-019-4549-4. PMID: 31615508; PMCID: PMC6794890.
20. Picker, D., Heard, K., Bailey, T.C. et al. The number of discharge medications predicts thirty-day hospital readmission: a cohort study. BMC Health Serv Res 15, 282 (2015). https://doi.org/10.1186/s12913-015-0950-9
21. Pereira F, Verloo H, Zhivko T, *et al* Risk of 30-day hospital readmission associated with medical conditions and drug regimens of polymedicated, older inpatients discharged home: a registry-based cohort study *BMJ Open* 2021;11:e052755. doi: 10.1136/bmjopen-2021-052755

ICCS Camera Ready Version 2022
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-08757-8_12