Machin Learning Approaches for WIES Value Prediction

MATH2191 - Applied Research Project

Adnan Hussain Qazi - s3701077
Alberto Romo - s3676768
Sakshi Mantri – s3701129

Western Health

RMIT UNIVERSITY

INTRODUCTION

The Victorian Government uses Weighted Inlier Equivalent Separation (WIES) value as their activity-based refunding policy for public hospitals. Each patient admitted to the hospital is assigned a WIES value which is determined by the classification of the patient into a Diagnosis Related Group (DRG). There are approximately 800 DRGs a patient can be classified to depending upon his/her illness and the treatment provided. At the end of the fiscal year hospitals are then paid for their total WEIS activity.

At present this process of classifying a patient to a particular DRG is carried out manually by clinical coders which takes around 20 days after the patient is discharged. The problem that this project attempts to solve revolves around this 20-day delay, since Western Health is also required to provide an estimate of year-to-date WIES revenue to the Department of Health and Human Services in the first week of each month. Accelerating the monthly WIES estimation process would allow Western Health to reduce uncertainty and improve financial decision making.

AIM

The aim of this study is to use the limited data that is available at the time of discharge, to predict the DRG code of a patient. An accurate predictive model will help minimizing the 20-day DRG delay period, and Western Health would be able to provide the Department of Health with accurate figures in a timely manner. The nature of DRG codes being 800 categories could make them extremely difficult to predict. Also it is not preferred to use machine learning algorithms for such large classification set.

So the second aim of the project is to directly predict WEIS (from readily available inpatient data and historical data). WEIS is a continuous variable, rather than categorical, and thus would allow the use of most forecasting algorithms and would be much more likely to be predicted accurately to a reasonable degree.

METHOD

To predict DRG, following classification algorithms were used -

- K-nearest Neighbor,
- Naïve Bayes and
- Decision tree

These were found to be approximately 50% accurate. Hence the focus was moved to predicting WIES Value directly for which various regression techniques were used.

Predicting WIES

- Data cleaning and preprocessing – Checked for missing or insane values.
- One hot encoding – To encode all categorical features into a set of numerical features.
- Scaling – Min-max scaling was performed to normalize all descriptive features.
- Feature Selection and Ranking – Ranked most important 10 features using Random Forest Importance to gain insight into the problem at hand.

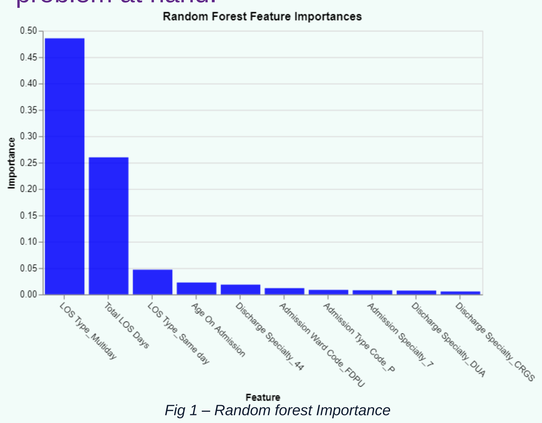

Fig 1 – Random forest Importance

- Test-train split – Data was split in 70:30 training and testing dataset.
- Modelling – Regression techniques were applied using Pipeline methodology.
  1. Random Forest
  2. Ridge Regression
  3. Neural Network
- Testing – To fit the model on test data 10-fold stratified cross validation technique was used.
- Future use – The model was saved and tested on a dummy data so as to be supplied to Western Health.

RESULTS

1. Random Forest

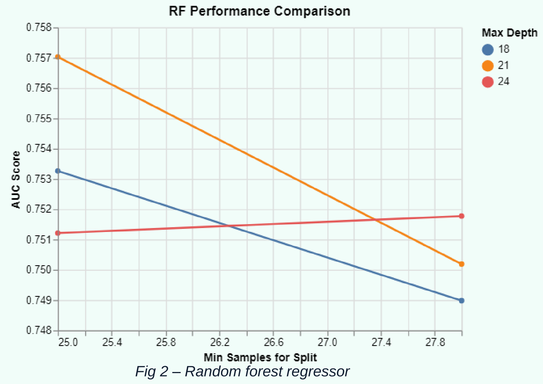

Fig 2 – Random forest regressor

Random Forest came out to be the best of three with an accuracy score(R^2) of 76%.

2. Ridge Regression

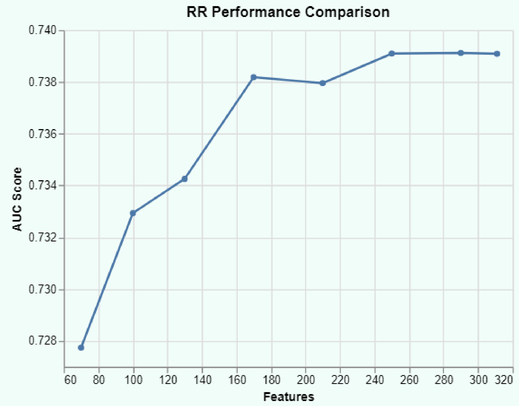

Fig 3 –Ridge regression

Accuracy score 74%

3. Neural Network

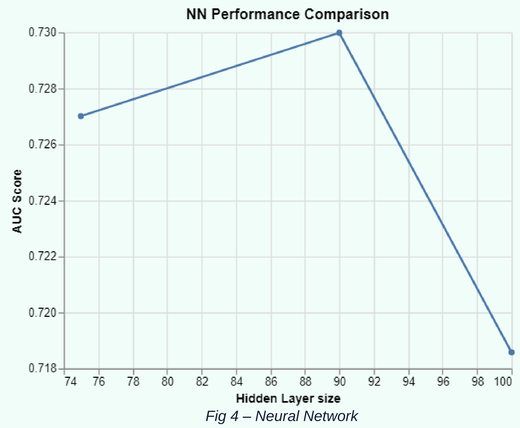

Fig 4 – Neural Network

Accuracy score 73%

If we see the actually required value i.e. sum of WIES for a particular month to be reported to Department of Health we get even better results. On testing the data with different dummy test datasets, WIES is predicted and when summarized the total is found to be 95-99% accurate.

CONCLUSIONS

- This project was successful in creating three machine learning regression models that could produce quality WIES estimates for Western Health's monthly reimbursement reporting to the Department of Health.
- Regression models for a continuous target variable (WIES) proved to be more plausible and precise than models for large multiclass classification (DRG codes).
- After several ML regression algorithms were tested the most precise (Random Forest Regressor) was handed to Western Health in the form of a Python file that can predict the sum of WIES in seconds by just uploading the new data and running the file.
- It is known that the Department of Health updates WIES values periodically, therefore it is recommended annually to retrain the model with the most recent data in order to get more accurate predictions.

REFERENCE

- Daniel Garten, R. K. (2015). Machine Learning Approaches for Early DRG Classification and Resource Allocation. INFORMS Journal on Computing, 718-734.
- Cournapeau, D. (2007). Nearest Neighbors. Retrieved from Scikit Learn: https://scikit-learn.org/stable/modules/neighbors.html
- Cournapeau, D. (2007). Naive Bayes. Retrieved from Scikit Learn: http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.46.1529

ACKNOWLEDGEMENTS

We would like to express our sincere gratitude to RMIT and our Industry supervisors for providing this invaluable project opportunity, offering industry knowledge and expert assistance in the field.