International Journal of Information, Control and Computer Sciences
ISSN: 2517-9942
Vol:6, No:2, 2012

Decision Trees for Predicting Risk of Mortality using Routinely Collected Data

Tessy Badriyah, Jim S. Briggs, and Dave R. Prytherch

***Abstract*—**It is well known that Logistic Regression is the gold standard method for predicting clinical outcome, especially predicting risk of mortality. In this paper, the Decision Tree method has been proposed to solve specific problems that commonly use Logistic Regression as a solution. The Biochemistry and Haematology Outcome Model (BHOM) dataset obtained from Portsmouth NHS Hospital from 1 January to 31 December 2001 was divided into four subsets. One subset of training data was used to generate a model, and the model obtained was then applied to three testing datasets. The performance of each model from both methods was then compared using calibration (the $\chi 2$ test or chi-test) and discrimination (area under ROC curve or c-index). The experiment presented that both methods have reasonable results in the case of the c-index. However, in some cases the calibration value ($\chi 2$) obtained quite a high result. After conducting experiments and investigating the advantages and disadvantages of each method, we can conclude that Decision Trees can be seen as a worthy alternative to Logistic Regression in the area of Data Mining.

***Keywords*—**Decision Trees, Logistic Regression, clinical outcome, risk of mortality.

I. Introduction

THE expected outcome of this study is to contribute to the building of effective and efficient methods to predict clinical outcomes for all general hospital admissions using routinely collected data, i.e. that which is available for the vast majority of patients admitted to a hospital, thus giving access to a "large" dataset. The particular outcome to be investigated is the risk of death ("mortality on discharge") as one of a number of possible adverse clinical outcomes.

Many studies [2, 5-11] confirm that Logistic Regression is the gold standard method to predict clinical outcome, especially to predict risk of death. However, a recent study[1] (Asiimwe, A. 2007) showed that Decision Trees as a Data Mining technique outperformed Logistic Regression, in particular when using a Chronic Obstructive Pulmonary Disease (COPD) dataset. So, in this research we decided to use Decision Trees as the primary method to compare with Logistic Regression.

Tessy Badriyah is a PhD student at the University of Portsmouth, United Kingdom. She is a lecturer from the Electronics Engineering Polytechnic Institute of Surabaya, Indonesia (email: tessy.badriyah@port.ac.uk or tessy@eepis-its.edu).

Jim S. Briggs is the Leader of The Centre for Healthcare Modelling and Informatics (CHMI) research group at the University of Portsmouth. He is now with the School of Computing, University of Portsmouth, United Kingdom (email : jim.briggs@port.ac.uk).

Dave R. Prytherch is a Senior Research Fellow in the Centre for Healthcare Modelling and Informatics at the University of Portsmouth (email: david.prytherch@gmail.com).

.

The main contribution of this paper is to show the feasibility of applying Decision Trees to predict clinical outcome, as well as providing a investigation into the advantages and disadvantages of using Decision Trees and Logistic Regression as a standard method commonly used in predicting clinical outcome, in this case the risk of mortality.

The rest of this paper is organised as follows: In Section 2, we explain the related work on predicting clinical outcomes using routinely collected data. In Section 3, we explain the datasets that have been used for the experiments, and explain our method of analysis used to assess the models. Section 4 presents the results of our experiments and discusses the results, and we conclude the paper and plan future work in Section 5.

II. Related Works

This research uses administrative and laboratory data which has been obtained from the hospital pathology and administrative computer systems at Portsmouth NHS Hospitals Trust.

This study focuses on predicting clinical outcome for all general admissions to a hospital using routinely collected data.

Generalised to all admissions, Prytherch et al. [8] have demonstrated the prediction of hospital outcome for general medical patients (i.e. including non-surgical cases) using routinely collected data. That study raised the possibility that the surveillance and treatment of patients might be categorised by early risk assessment in the future. High-risk patients could then get intensive care and, in the case of low-risk patients, it might even be possible to safely send them home.

The pathology data items used were those from the first routinely collected haematology and biochemistry blood tests, i.e. haemoglobin, white cell count, and serum levels of urea, albumin, creatinine, sodium and potassium. The administration data items extracted were patient age at admission, patient sex, mode of admission (elective or emergency) and outcome (survival or non-survival) at hospital discharge. A model was built using a training set (Q1) corresponding to three months' worth of patients. Application of the model to the validation sets produced c-indices of 0.779 (Q2), 0.764 (Q3) and 0.757 (Q4), respectively, indicating good discrimination, and also gave $\chi 2 = 9.43$ (Q2), $\chi 2 = 7.39$ (Q3) and $\chi 2 = 8.00$ (Q4) (p-values of 0.307, 0.495 and 0.433) for 8 degrees of freedom, indicating good calibration.

III. Methods

A. Data Description

This research uses administrative and laboratory data which has been obtained from the hospital pathology and administrative computer systems at Portsmouth NHS

Hospitals Trust. This particular dataset was the Biochemistry and Haematology Outcome Model (BHOM) dataset, which contains 9497 adult hospital discharges, and it was divided into four subsets, one for data training and three for data testing. Training data (data from 1 January to 31 March 2001 (Q1) – n1 = 2257) was used to generate a model. The model obtained was then applied to three testing data sets (1 April - 30 June (Q2) n2 = 2335, 1 July - 31 September (Q3) n3 = 2361, 1 October - 31 December 2001 (Q4) n4 = 2544).

The fields in the dataset are : Death at discharge - F=alive, T=dead (class attribute), Age at admission, Mode of admission – (emergency or elective), Gender, Haemoglobin, White cell count, Urea, Serum sodium, Serum potassium, Creatinine and Urea creatinine.

B. Method of Analysis

The statistical analyses used to assess the overall performance of the model are calibration and discrimination. Calibration (or reliability) is the accuracy of risk predictions and refers to whether the predicted probabilities agree with observed probabilities. Calibration is most suited to a problem where we would like to predict risk in the future. This is because calibration measures how well the predicted probabilities correctly estimate a future event.

$$\chi^2(i) = \sum_{i=1}^{n} \frac{(reported(i) - predicted(i))^2}{predicted(i)} \quad (1)$$

Equation (1) is the formula to calculate the calibration using $\chi 2$ test (chi-test) value. Individual records in the validation subset are grouped by risk range. For each risk, the predicted number of deaths is compared to the number observed. Goodness-of-fit is assessed using the $\chi 2$ test (chi-test). As this is a null hypothesis test, p values less than 0.05 indicate evidence of significant lack of fit.

Discrimination is the ability to correctly discriminate between two conditions, in this case, between survivor and non-survivor. The discriminant ability of the models is assessed using receiver-operating characteristics (ROC) curves. The area under the ROC curve, summarised by the c-index, can range from 0.5 (no predictive ability) to 1 (perfect discrimination). Reasonable discrimination is indicated by c-index values of 0.7-0.8 and good discrimination by values exceeding 0.8.

IV. Results

We followed the research that has been undertaken by replicating the results reported in Prytherch et. al. [8][11], . We took the same path by using the same data, and then used the same method (Logistic Regression) to generate a model, and we also used the same analysis method (discrimination and calibration).

Using the same dataset, Decision Trees were then used to generate a model. Finally we could compare the performance of the Logistic Regression Model and Decision Trees model when applied to the testing data.

We used SPSS software to generate the Logistic Regression and Decision Trees models, and also developed code in Matlab to provide stratified modelling.

A. Logistic Regression Model

Logistic regression using SPSS tools produced the following outcome model based on the BHOM Q1 training set :

Ln(R/(1-R))= -23.194+(-0.013 × gender)+((18.714 × mode of admission)+ (0.053 × age on admission)+ (0.018 × urea)+(-0.001 × Na+)+ (-0.101 × K+)+(-0.047 × albumin)+ (-0.037 × haemoglobin)+ (0.067 × white cell count)+ (0.001 × creatinine)+ (2.744 ×urea/creatinine). (2)

This model concurs with that stated in [8] and validates our algorithm.

By using Q1 as training data and SPSS tools to generate a Logistic Regression model and then applying it to test data Q2, we obtained the results of stratified modelling shown in Table (1).

As seen in Table (1), in order to measure the calibration, we stratified the risk values into several levels from the lowest level ($0 \leq$ risk < 5) up to the highest risk band level ($50 \leq$ risk ≤ 100). For each band, we then calculated the total of mean predicted risk, the total number of deaths predicted, the total number of deaths reported and the value of χ^2 (chi-test).

TABLE I
LOGISTIC REGRESSION USING SPSS GOODNESS-OF-FIT BY HOSMER-LEMESHOW X2 STATISTIC FOR (Q2) DATA COVERING PERIOD 1 APRIL–-30 JUNE 2001

Risk bands	No. of cases	Mean predicted risk (%)	Predicted deaths	Reported deaths	$\chi 2$
≥ 0 to < 5	1037	2.07	22	16	1.44
≥ 5 to < 7.5	298	6.21	18	17	0.13
≥ 7.5 to <10	240	8.65	21	22	0.08
≥ 10 to <12.5	202	11.14	22	27	1.07
≥ 12.5 to < 15	150	13.60	20	20	0.01
≥ 15 to < 20	174	17.22	30	31	0.03
≥ 20 to < 25	97	22.18	22	22	0.01
≥ 25 to < 33	77	28.09	22	12	5.96
≥ 33 to < 50	46	39.10	18	17	0.09
≥ 50 to ≤ 100	14	61.00	9	7	0.71
≥ 0 to ≤ 100	2335	8.71	203	191	9.53

Calibration: $\chi 2 = 9.53$; 8 d.f.; p-value = 0.483;
discrimination : c-index = 0.779.

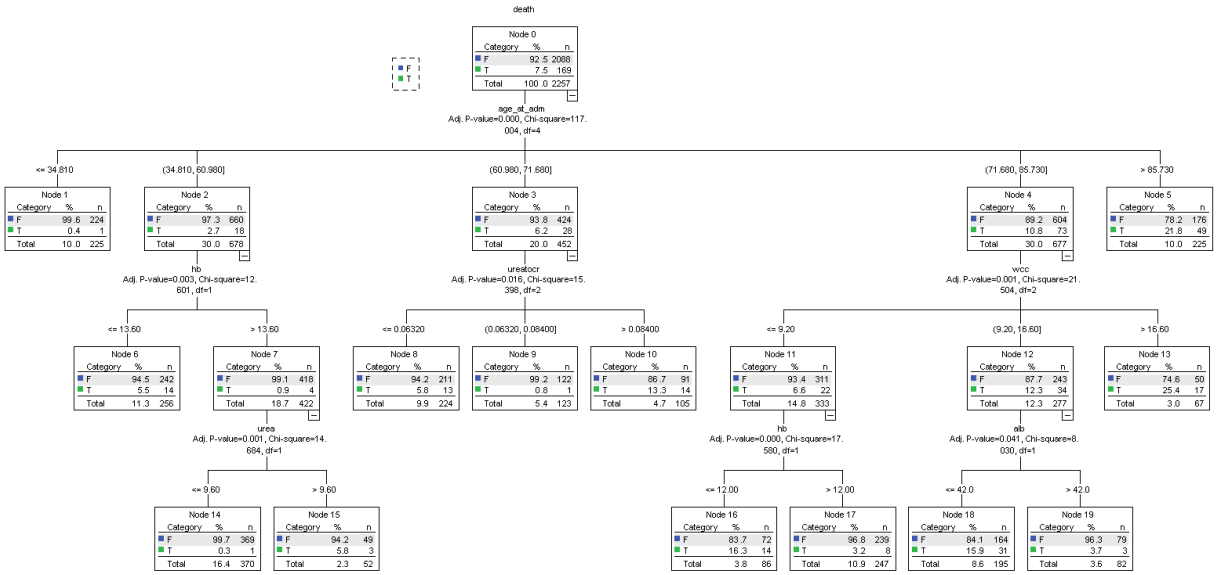

Fig. 1 Decision Trees Model for BHOM dataset

B. Decision Trees Model

We choose to use the CHAID method in SPSS. In each step, CHAID always chooses the independent variable that has the strongest relation with the dependent variable. This was very relevant to the problem to be solved. Figure (1) shows the resulting Decision Trees model.

The number of risk bands in Decision Trees is determined by the number of terminal nodes (leafs) that exist on the tree. Based on the modelling results in Figure (1), we can see that there are as many as 13 risk bands.

The lowest level of risk band is node 14, the probability of risk of mortality of only 0.3%: only one person is reported dead from a total of 370 people in this node. The percentage of people who fall into this node is 16.4% of the total number of patients. Whereas if we look at the highest level of risk band that is at node 13, the probability of risk of mortality is 25.4%: 17 people are reported dead out of a total of 50 people in this node.

For our Decision Tree model (applied to Q2 test data), we obtained the results of stratified modelling as shown in Table (2).

This did not produce good calibration; in other words there is a significant lack of fit as indicated by the value of χ^2 (chi-test) is 67.05 (p-value < 0.05). This is caused mainly by the failure of decision trees to predict risk of mortality at node 9 when risk band = 3, the mean predicted risk = 0.8 with a value of χ^2 (chi-test) = 36.34; and also at node 17 when risk band = 4, the mean predicted risk = 3.2 with a value of χ^2 (chi-test) = 10.64. The discrimination value obtained, however, is 0.735, which indicates reasonable discrimination, although it is still smaller than the discrimination value obtained by Logistic Regression (0.779).

TABLE II
LOGISTIC REGRESSION USING SPSS GOODNESS-OF-FIT BY HOSMER-LEMESHOW X2 STATISTIC FOR (Q2) DATA COVERING PERIOD 1 APRIL–-30 JUNE 2001

Risk bands	No. of cases	Mean predicted risk (%)	Predicted deaths	Reported deaths	χ2
1	334	0.3	1	2	1.34
2	221	0.4	1	1	0.00
3	155	0.8	1	8	36.34
4	226	3.2	7	16	10.64
5	99	3.7	4	9	8.29
6	270	5.5	15	16	0.11
7	55	5.8	3	2	0.46
8	223	5.8	13	17	1.35
9	99	13.3	13	8	2.36
10	249	15.9	40	28	4.03
11	86	16.3	14	12	0.34
12	233	21.8	51	46	0.57
13	85	25.4	22	26	1.22
All	2335	7.9	184	191	67.05

Calibration: χ2 = 67.05; 11 d.f.; H–L p-value = 0.303434;
Discrimination : c-index = 0.735.

C. Comparison between the two methods

Table (3) compares the performance between Decision Trees and Logistic Regression in the case of discrimination (c-index) and calibration (χ2).

Table (3) shows that for all testing data, the Decision Tree model has a significant lack of fit. This is indicated where χ2 of Q2, Q3 and Q4 obtained a quite high result (67.05; 159.35; 133.07), even though, in the stratified model as shown in table

TABLE III
COMPARISON MEAN PREDICTED RISK AND DISCRIMINATION BETWEEN LOGISTIC REGRESSION AND DECISION TREES

Dataset	No. of cases	Logistic Regression		Decision Trees	
		c-index	$\chi 2$	c-index	$\chi 2$
Q2	2335	0.779	9.53	0.735	67.05
Q3	2361	0.764	23.55	0.721	159.35
Q4	2544	0.757	6.66	0.700	133.07

(2), this was caused only by a small number of nodes, while overall it can be said that almost all risk bands in Decision Trees have the ability to discriminate with similar values between predicted and reported.

Logistic Regression outperformed Decision Trees for all testing datasets (Q2, Q3 and Q4) in the case of discrimination..

The discrimination (c-index) value is exactly the same as Prytherch, et. al [8], however the calibration value is rather different on this experiment where $\chi 2$ of Q2 obtained a quite high result (23.36), therefore the p-value 0.0002932 indicates evidence of lack of fit.

V. CONCLUSIONS AND FUTURE WORK

Many studies [2, 5-11] confirm that Logistic Regression is the gold standard method to predict clinical outcome, especially to predict risk of death. In this paper, the Decision Trees model has been proposed to solve specific problems that commonly use Logistic Regression as a solution. The Logistic Regression model provides a constant value for each attribute (Equation 2). This kind of model is like a "black box", where the most influential attribute is unknown. On the contrary, in the Decision Trees model (Figure 1), it can be seen that the age_at_adm (age at admission) attribute is at the highest level (root) of the tree. So, it is clear that patient's age is the most influential on clinical outcome to predict risk of mortality. This make sense, because elderly people are more likely to die and young people are more likely to have a quick recovery. When we can start with the root of the tree, we can continue to considering the other attributes below the root.The advantage of Decision Trees is that the resulting model can be interpreted by humans as decision rules. In other words, this method has the advantage of human interpretability of the results.From our experiments, we can conclude that Logistic Regression and Decision Trees are both effective means of constructing models to predict risk of mortality. Both methods provided reasonable discrimination.The experiment conducted in this paper did not optimise the parameters in the Decision Trees. It is not known yet whether Decision Trees can outperform Logistic Regression when appropriately parameterised. Although when applied to all testing data, Logistic Regression outperformed Decision Trees, considering the advantages belong to Decision Trees, we can conclude that the Decision Trees can be seen as a worthy alternative to Logistic Regression in the area of Data Mining.

For future planned research, in addition to optimising the parameters in the Decision Trees, we also want to try other methods in Data Mining, including Support Vector Machines (SVM), Radial Basis Networks (RBN), K-Nearest Neighbours (KNN) and others.

REFERENCES

[1] Asiimwe, A. (2007). Morbidity and mortality in patients with stable and unstable COPD: Construction and validation of a prediction model using routinely collected data, University of Portsmouth.
[2] B. Silke, J. Kellett, T. Rooney, K. Bennett, and D. O'Riordan (2010). An improved medical admissions risk system using multivariable fractional polynomial logistic regression modeling, QJM 103(1): 23-32 doi:10.1093/qjmed/hcp149
[3] Cook, N. R. (2008). Statistical evaluation of prognostic versus diagnostic models: Beyond the ROC curve. Clinical Chemistry, 54(1), 17-23. doi: 10.1373/clinchem.2007.096529
[4] Copeland, G.P., D. Jones and M. Walters (1991). POSSUM: a scoring system for surgical audit. Br J Surg 78(3): p. 355-60.
[5] Pine, M., B. Jones, and Y.-B. Lou (1998). Laboratory values improve predictions of hospital mortality, Int J Qual Health Care 10(6): 491-501 doi:10.1093/intqhc/10.6.491
[6] Prytherch, D. R., J.S. Briggs, P.C. Weaver,, P. Schmidt, & G.B. Smith, (2005). Measuring clinical performance using routinely collected clinical data. Medical Informatics and the Internet in Medicine, 30(2), 151-156. doi: 10.1080/14639230500298966
[7] Prytherch, D. R., B.M.F. Ridler, S. Ashley & Audit Res Comm Vascular Soc (2005). Risk-adjusted predictive models of mortality after index arterial operations using a minimal data set. Br J Surg, 92(6), 714-718. doi: 10.1002/bjs.4965
[8] Prytherch, D. R., J.S. Sirl, P. Schmidt, P.I. Featherstone, P.C. Weaver, & G.B. Smith (2005). The use of routine laboratory data to predict in-hospital death in medical admissions. Resuscitation, 66(2), 203-207. doi: 10.1016/j.resuscitation.2005.02.011
[9] Prytherch, D. R., G.B. Smith., P.E. Schmidt & P.I. Featherstone (2010). ViEWS-Towards a national early warning score for detecting adult inpatient deterioration. Resuscitation, 81(8), 932-937. doi: 10.1016/j.resuscitation.2010.04.014
[10] Prytherch, D. R., M.S. Whiteley, B, Higgins, P.C. Weaver,,W.G. Prout, & S.J. Powell (1998). POSSUM and Portsmouth: POSSUM for predicting mortality. Br J Surg, 85(9), 1217-1220.
[11] Tang, T., S.R. Walsh,, D.R. Prytherch, T. Lees, K. Varty, J.R. Boyle & Assoc Res Comm Vascular Soc. (2007). VBHOM, a data economic model for predicting the outcome after open abdominal aortic aneurysm surgery. Br J Surg, 94(6), 717-721. doi: 10.1002/bjs.5808