Cost Sensitive Classification in Healthcare

ST309 GROUP PROJECT REPORT

CANDIDATE NUMBER: [redacted] (50%), [redacted] (50%)

1. INTRODUCTION

Hypothyroidism and its diagnosis

Hypothyroidism is a disorder of the body's endocrine system, in which the thyroid gland does not produce enough thyroid hormone (Weetman 1997). If untreated, it can result in symptoms such as tiredness, constipation, weight gain, muscle or joint pain, and depression. In rare but extreme cases, severe untreated hypothyroidism may lead to Myxedema, a life-threatening coma (Weetman 1997). On the other hand, mis-diagnosing a healthy patient as hypothyroid is also costly. The healthy patient may be administered treatment to promote thyroid function, which could in turn result in hyperthyroidism. As such, it is an important problem to determine whether a patient referred to the clinic is hypothyroid.

In the prevailing medical practice, doctors diagnose if a patient referred to a clinic is hypothyroid based on a set of observable attributes and five separate medical tests. The observable attributes include the patient's age, sex, symptoms of illness, and whether they are currently taking certain medications etc. The five medical tests, TSH, TT4, T4U, T3, TBH, measure hormone levels within the patient's bloodstream (Cooper 2001). With this information, doctors determine if a patient suffers from hypothyroidism.

Predictive analytics in healthcare

The challenge of determining if a patient is hypothyroid can be cast as a classification problem in the field of data analytics. We aim to use historical data (i.e. training data) to identify the characteristics of normal patients and hypothyroid patients. We then predict the class of new patient data (i.e. test data) based on the identified characteristics (James et al 2013). In this way, our project contributes to the emerging field of predictive healthcare analytics.

Indeed, the use of such predictive data analytics has entered healthcare policy in the past few years (Janke et al 2016). For example, a cardiac research program at California Pacific Medical Center in San Francisco, USA has initiated numerous projects to analyze huge amounts of medical data on patients with cardiovascular diseases to develop sophisticated cardiac risk models (Raghupathi 2013). Additionally, the Seton Healthcare hospital in Texas, USA has also applied data analytics to predict readmission rates for patients with congestive heart failure. Using data from physician notes and discharge summaries, Seton engaged in predictive analytics to identify trends and patterns in patient care and outcomes by detecting correlations or disparities (Raghupathi 2013).

Our project also makes an important contribution in the field of predictive healthcare analytics in terms of *test-cost sensitive* classification. While the current healthcare system runs all five medical tests on each patient, our hypothesis is that some of these tests are unnecessary. We suggest that we may achieve the same misclassification costs but with fewer tests and thus lower costs. Hence, the problem is classifying if a patient is hypothyroid, while taking into account both the misclassification costs as well as the costs of each test. We aim to construct a classification model which minimises the sum of the costs of misclassification and the costs incurred in conducting each test. This is a significant innovation on the status quo.

2. DATA

Description of data

We sourced a dataset containing 7200 instances with 21 attributes, obtained from the Garvan Institute of Medical Research based in Sydney, Australia (Quinlan 1987) on the UCI Machine Learning Repository. Of the 21 attributes in the data, 15 are binary and 6 are continuous. The binary attributes are 'sex', 'on thyroxine', 'query on thyroxine', 'on antithyroid medication', 'sick', 'pregnant', 'thyroid surgery', 'I131 treatment', 'query hypothyroid', 'query hyperthyroid', 'lithium', 'goitre', 'tumour', 'psych'. The conditions variables are 'age' and the reading on the 5 tests, 'TSH', 'T3', 'TT4', T4U', 'TBG'.

There are several inherent issues with this dataset. First, the dataset contains 368 'positive' (i.e. hypothyroid) cases and 6832 'negative' cases. The true positive rate is thus 5.1%, meaning that the dataset is imbalanced. This may mean that the information on hypothyroid patients may be overwhelmed by that on non-hypothyroid patients such that the signal on 'positives' is too weak to be picked up in supervised machine learning (Provost and Fawcett 2013). Over-sampling and bagging techniques may be required to account for this (James et al 2013). Second, the dataset provided also does not contain variables on common observable symptoms of hypothyroidism, such as tiredness, constipation, weight gain, muscle and joint pain and increased sensitivity to cold. This means that our classification models are trained on less information than would be available to medical doctors. This should be kept in consideration when comparing the misclassification rates of our models against the baseline misclassification rates of medical doctors.

Data preparation

We divided the dataset into 5200 instances (4932 negative, 268 positive) for training data and the remaining 2000 instances for test data (1900 negative, 100 positive). Notably, the dataset originally split the patients into three classes - 'hypothyroid', 'healthy' and 'hyperthyroid'. We reclassified the data into 'hypothyroid' and 'normal', the latter class created by combining the 'healthy' and 'hyperthyroid' classes. This was done to match the outcome variables to the problem statement - to diagnose if a patient suffers from hypothyroidism. Such a reclassification also facilitated the decision tree classification model with a cut-off probability. However, the reclassification has a disadvantage in that we do have an medical interpretation of the 'normal' cases, as they refer to both healthy and hyperthyroid patients. In any case, the number of hyperthyroid cases is less than 3%.

Only minor data cleansing and transformation was conducted. The age variable in the dataset was denoted in decimal places (i.e. 0.52 indicates 52 years of age). We scaled this variable by a factor of 100, to reflect the age more intuitively in years. No transformation was conducted on the remaining continuous variables, the patient's results on the five medical tests, to preserve the numeric integrity of the test scores (Provost and Fawcett 2013). The rest of the variables are binary and require no transformation.

There are costs incurred when using the medical test ('TSH', 'T3', 'TT4', 'T4U', 'TBG'). The dataset used contained the test scores of all the medical test apart from 'TBG'. Hence, we estimated the cost of the 'TBG' test by comparing the known test costs to current prices provided by a medical lab online: Accessalabs.

Medical Tests	Cost per person
TSH	22.78
T3	11.41
TT4	14.51
T4U	11.41
TBG	41.28 (estimated)

Medical Test Attributes and associated cost

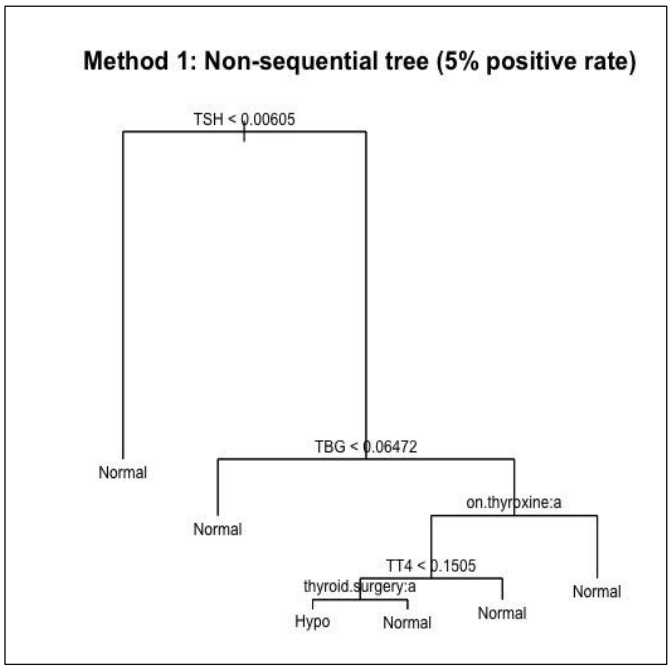
3. METHODS AND RESULTS

The current healthcare system runs all 5 tests simultaneously on all patients to reach a diagnosis. Our hypothesis is that we may achieve the same misclassification costs but with fewer tests and thus lower costs. To achieve this, we develop four classification methods, labelled Method 1-4.

We evaluate the misclassification rate and overall tests costs for each Method. Since this is a healthcare problem and patient's lives are at stake, misclassification of cases is viewed very severely. We calculate the total medical test costs associated with classifying our test dataset of 2000 instances. We calculate the test cost of this decision tree by tracking the medical tests each patient had to undergo and then summing the medical test costs over the entire test dataset. For example, if a patient was classified as 'Normal' at a given terminal node, the test cost incurred by the patient will be the sum of cost of all the medical tests he/she had previously undergone to reach that terminal node.

Method 1: Simple Classification Tree

We first build a simple classification decision tree on the entire training dataset (5200 instances) using all the variables available. The positive rate in this dataset is 5%.



Method 1: A simple classification tree with 5% positive rate training data

Misclassification costs - The misclassification costs in this model are balanced. A false negative, classifying a patient with hypothyroidism as 'normal' will mean that the patient is denied treatment, leading their quality of life to be severely diminished. On the other hand, a false positive, classifying a 'normal' patient as 'hypothyroid' will mean the healthy patient is administered treatment to promote thyroid function, which could in turn result in hyperthyroidism. This also diminishes the patient's quality of life. Hence, we index the same misclassification costs to both types of misdiagnosis.

Cost Benefit Matrix	Actual		
Predicted		Hypothyroid	Normal
	Hypothyroid	0	-15
	Normal	-15	0

Cost Benefit Matrix 1: Misclassification cost for overall classification

Cross Validation - To avoid overfitting, when a model is tailored too much to the training data at the cost of generalization to previously unseen data points, we further performed 10-fold cross validation and pruned the decision tree accordingly. This is done by finding the finding the number of nodes with the associated decision tree that has the least deviation (James et al 2013). As a result, Method 1 is a simple classification tree that is pruned to counter overfitting of data.

Result 1: Simple Classification Tree

From the classification decision tree above, we observe that the largest information gain came from the tests of 'TSH' and 'TBG'. This is unsurprising as the medical tests are expected to provide a more accurate diagnosis of hypothyroidism than the observable characteristics. Furthermore, as expected, using a decision tree which includes the medical test data, we are able to classify patients from our test data very accurately. The misclassification rate of this tree is only 0.1%, or 2 out of 2000 test patients.

Method 1	Actual		
Predicted		Hypothyroid	Normal
	Hypothyroid	100	2
	Normal	0	1898

Method 1 Results: Classification on test data (2000 instances)

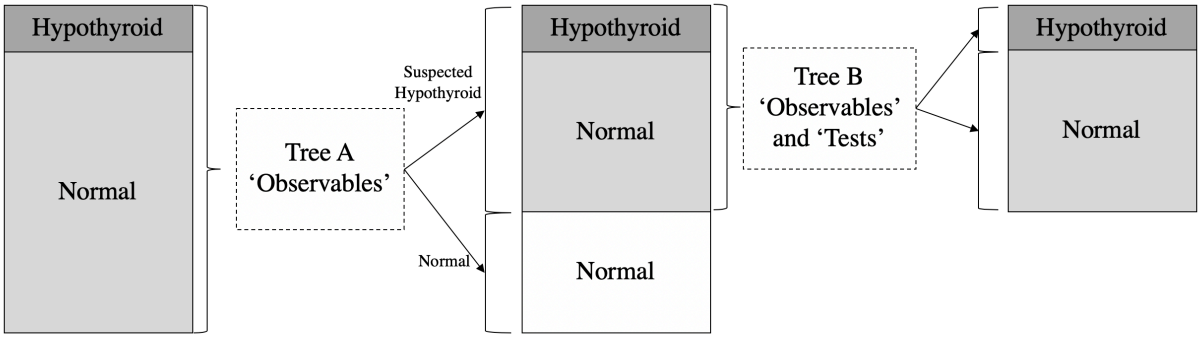
In terms of test cost however, this model is not ideal. Given the significant information gain from the medical tests, the model requires all patients to take the 'TSH' test first, while some patients may also be subjected to the 'TBG' and 'TT4' tests further down the line. As a result, this classification model involves a relatively high number of medical tests and incurs high medical test costs of $55,058.58 for our entire test dataset.

Method 2: 2-Step Sequential Tree

To improve on Method 1, the simple classification tree, we propose a 2-step sequential tree to try to achieve the same misclassification rate but at lower costs. The intuition behind this model is simple: Based solely on the costless observable characteristics, we believe that we can classify a group of patients as ‘normal’ without requiring any medical tests. Hence, we are able to save on the medical test costs for this group of patients.

Cast as a data analytics problem, this is a 2-step classification approach with 2 decision trees. The first tree (referred to as 'tree A') is only based on 'observable' attributes (such as 'age', 'sex', 'on thyroxine' etc.) which are costless. From this tree, we classify patients into 2 different categories of 'suspected hypothyroid' patients and 'normal' patients. Patients classified as ‘normal’ will be deemed as non-hypothyroid and will not proceed on to further tests. Patients classified as 'suspected hypothyroid' will receive further medical tests in order to make a diagnosis. As such, they will be moved to the second tree (referred to as 'Tree B'), where they are classified as either 'hypothyroid' or ‘normal’ based on all variables including the medical tests.

This two-step classification model thus achieves cost-savings as Tree A can costlessly identify a group of ‘normal’ patients, which can be correctly classified without having to take a single medical test.



Flow of 2-Step Model: ‘Tree A’ to classify patients as ‘Normal’ and ‘Hypothyroid’ (patients at risk) and ‘Tree B’ to make the final classification through the use of medical tests in order to accurately diagnose patients who have been deemed “at risk of hypothyroidism” by Tree A.

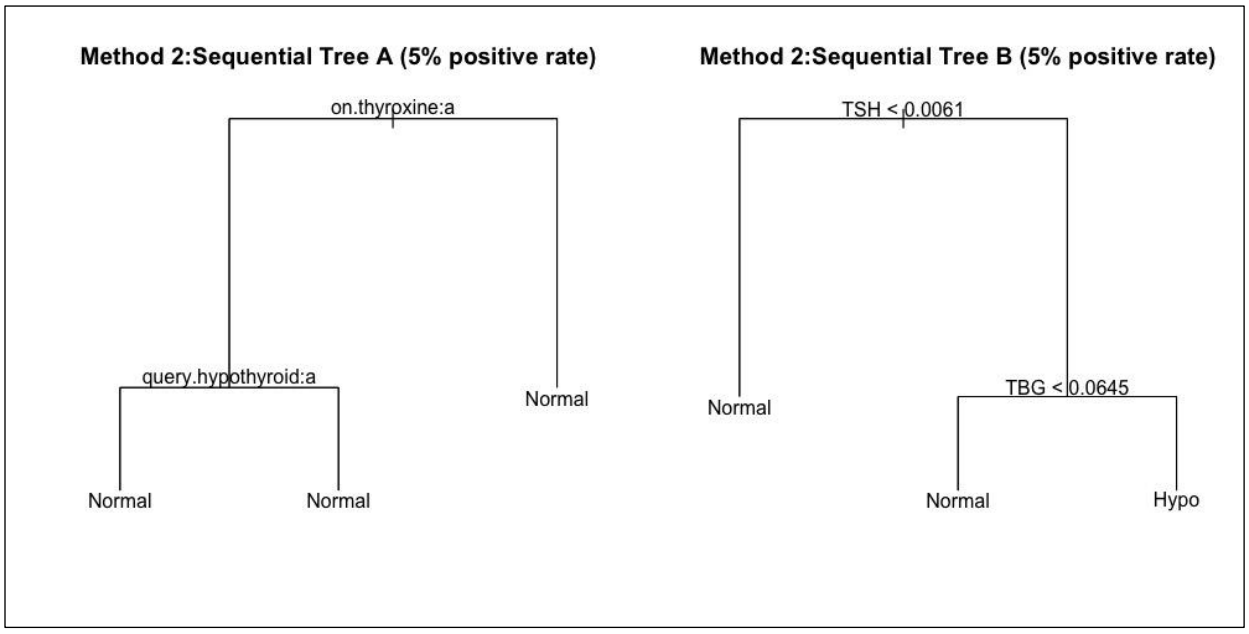
Misclassification costs - In the new 2-step sequential tree, we note that the misclassification costs for Tree A are now asymmetric. A false negative in Tree A, classifying a patient with hypothyroidism as normal, would mean that the patient is not subjected to further medical tests in Tree B. The hypothyroid patient will thus be undiagnosed and denied treatment, leading their quality of life to be severely diminished. The

misclassification cost of a false negative is thus very high. On the other hand, a false positive in Tree A, classifying a patient as hypothyroid when he is actually healthy, means that the patient would be moved to Tree B of the decision tree and subjected to additional medical tests. Since the patient is actually healthy, these additional medical tests are unnecessary. Hence, the misclassification cost of false positives is the test costs of the five medical tests incurred.

Cost Benefit Matrix	Actual		
Predicted		Hypothyroid	Normal
	Hypothyroid	0	-1
	Normal	-15	0

Cost Benefit Matrix 2: Misclassification cost for Tree A – Part A

We index the misclassification cost of false positives as '-1'. This value is associated with a real-world dollar amount, which is the average test cost that will be incurred to classify a patient. To emphasise the relative severity of false negatives compared to false positives, we denote the misclassification cost of false negatives as '-15'. We find the optimal cut-off probability for Tree A to be lower than that of 0.5 (default) due to the classification being more cautious against classifying someone as 'hypothyroid' when they are in fact 'normal'. In Tree B, the classification is the same as in Method 1, the simple classification model. Hence, as before, the misclassification costs in Tree B are balanced. Once again, we also conduct cross-validation to enhance the generalization of our model.



Method 2: 2-Step Sequential Tree consisting of Part A and Part B

Results 2: 2-Step Sequential Tree

Method 2 Part A		Actual	
		Hypothyroid	Normal
Predicted	Hypothyroid	13	81
	Normal	87	1819

Method 2 Results (Part A): Classification on test data (2000 instances)

Method 2 Part B		Actual	
		Hypothyroid	Normal
Predicted	Hypothyroid	13	0
	Normal	0	81

Method 2 Results (Part B): Classification on 'Hypothyroid' patients from Part A (94 instances)

Method 2: Final	Actual		
		Hypothyroid	Normal
Predicted	Hypothyroid	13	0
	Normal	87	1900

Method 2 Results: Final Classification on all test data (2000 instances)

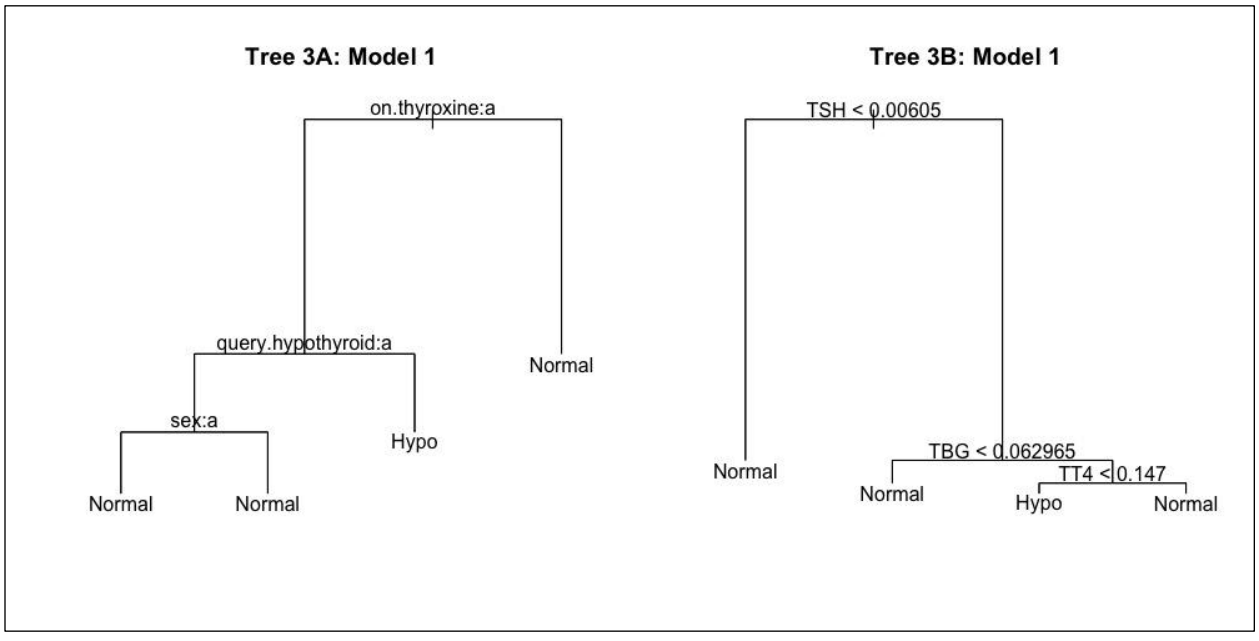
Method 2 has a far higher overall misclassification rate of 4.35%. From the 3 misclassification tables, we can see that the poor performance of Method 2 overall is due more specifically to Tree A. Tree A has a misclassification rate of 8.4%. To put this in context, the true positive rate is only 5.1%, meaning that if a decision tree had simply classified all patients as 'normal', it would obtain a lower misclassification rate than Tree A. By contrast, Tree B correctly classified all 94 patients it encountered. Since far fewer patients were sent to Step 2 (Tree B), it is unsurprising that the total medical costs fell while the misclassification rate increased significantly: the overall costs of Method 2 is $7,730.14, which is a 85% reduction on the costs of Method 1.

We identify that the problem with Tree A arises due to the imbalanced class distribution (there are only 5% positive cases). As Tree A is trained only on the observable variables, which do not convey much information about the patient's condition, the information on hypothyroid patients is overwhelmed by that on non-hypothyroid patients such that the signal on 'positives' is too weak to be picked up by the decision tree. As such, Tree A does poorly in identifying the hypothyroid cases. Once a patient is classified as 'normal' by Tree A, he/she does not move on to Tree B to undergo further medical tests. Thus, the poor performance of Tree A results in the poor performance of the model overall.

Method 3: 2-Step Sequential Tree (25% Bagged)

To improve on Method 2, which suffers from imbalanced class distribution, we increase the proportion of positive cases to 25% in the training data for Tree A and Tree B. Method 3 is thus a 2-step sequential Tree but with a balanced dataset for Tree A and Tree B.

Bagging - We create a more balanced training dataset with 25% positive cases. However, this training dataset is much smaller as we are omitted many negative cases. In order to train the model on more data, but maintain a 25% positive rate, we used the statistical technique bagging. To do this, we sample 10 different training sub-datasets - each containing 75% of 'normal' patients and 25% of 'hypothyroid' patients from the training dataset containing 5200 instances. This is done by using all 'hypothyroid' patients in each dataset and sampling a set of 'normal' patients without replacement. In this way, we are able to use a wider range of training data despite the increase in positive rates. We conduct bagging without replacement, to reduce correlation between the bagged trees. An example of a tree output is shown below (other 9 models in Appendix):



Method 3: 2-Step Sequential Tree consisting of Part A and Part B – One of 10 bagged models

Similar to the previous Methods, we also find the optimal cut-off probability under Method 3 taking into account the asymmetric misclassification costs of false negatives and false positives in Tree A. We also conduct cross-validation on 'Tree B' to prevent overfitting of the models.

To arrive at a final prediction, we combine the classification of the 10 tree models produced (each consisting of Tree A and Tree B) to make a majority aggregation on the final classification based on each individuals' tree optimal cut-off probability.

Result 3: 2-Step Sequential Tree (25% Bagged)

Method 3 has a misclassification rate of 1.45%. From the misclassification table, we see that Tree A correctly identifies 13.2% of 'normal' patients, saving the medical test cost on this group. Tree A is also cautious, achieving zero false negatives while having a high false positive rate. This is ideal, as the goal of Tree A is to correctly identify as many 'normal' cases as possible, without generating any false negatives. Tree B has a misclassification rate of 1.65%.

Method 3 Part A		Actual	
		Hypothyroid	Normal
Predicted	Hypothyroid	100	1649
	Normal	0	251

Method 3 Results (Part A): Classification on test data (2000 instances)

Method 3: Part B		Actual	
		Hypothyroid	Normal
Predicted	Hypothyroid	100	29
	Normal	0	1620

Method 3 Results (Part B): Classification on 'Hypothyroid' patients from Part A (1749 instances)

Method 3: Final Classification		Actual	
		Hypothyroid	Normal
Predicted	Hypothyroid	100	29
	Normal	0	1871

Method 3 Results: Final Classification on all test data (2000 instances)

Calculating the test costs associated with this method is challenging as we have 10 different 'Tree B' models that could put the same patient through different number of tests. However, since Tree A is able to correctly identify 12.55% of the overall number of patients, and hence save on the medical test costs on this group on at least the first medical test 'TSH', we estimate at least a 10.38% decrease in the overall medical cost compared to Method 1. The maximum estimated test cost of Method 3 is thus $49,430,80.

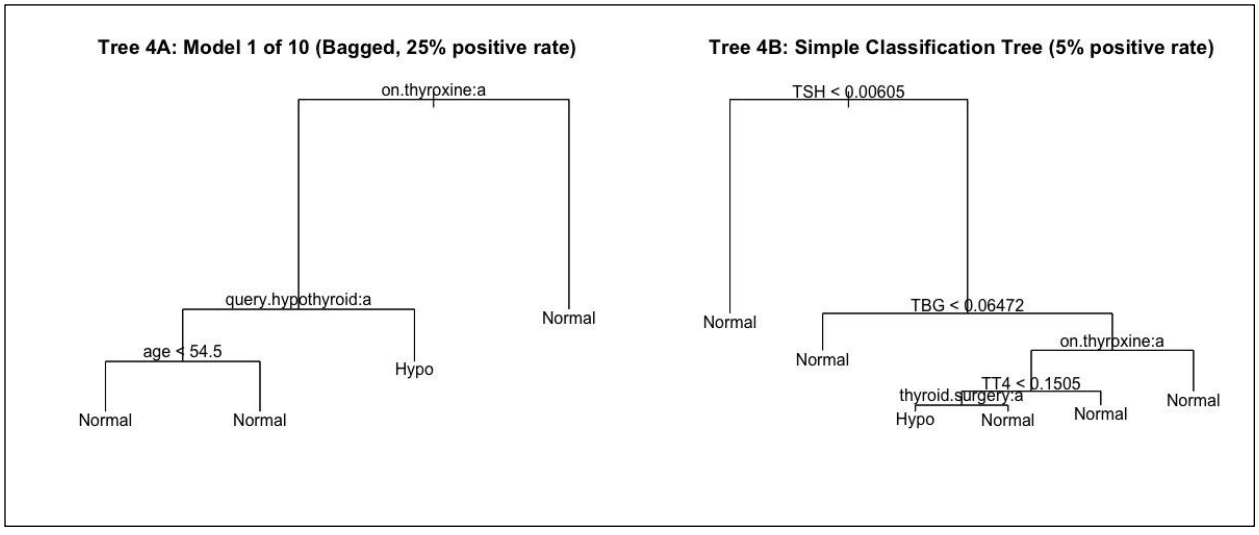
Analysing the results, we note that the misclassification rate of Method 3 is lower than in Method 2 (1.45% < 4.35%). This is due to the better accuracy of Tree A in Method 3. However, the overall classification using Method 3 is worse than Method 1. Even though both Tree B and Method 1 are trained on all the variables (including the medical test data), Tree B has a higher misclassification rate (1.65% > 0.1%). We pinpoint that this is because Tree B is trained on a training data with a 25% positive rate, but the testing data has a positive rate of only 5%. This means that the optimum cut-off probability based on the training data is too low.

Method 4: 2-Step Sequential Tree (Tree A - 25% Bagged, Tree B - 5% Simple Classification)

To recap, we have seen that a 2-step sequential tree can lower test cost by correctly classifying a group of patients as 'normal' without requiring any medical tests. In Method 4, we build on the results from Methods 1-3 to create an optimal 2-step sequential tree with the lowest misclassification rate.

By comparing Tree A in Method 3 and Method 2, we see that training the model on a balanced training data with 25% positive rate leads to a lower misclassification rate. This is due to the increased sensitivity of the fitted model to the signals from the positive cases. By comparing Tree B in Method 3 and the simple classification tree in Method 1, we observe that training the model on an imbalanced training data of 5% positive rate leads to a lower misclassification rate. This is to avoid the upward biased optimum cut-off probability from the training data.

Hence, our final proposed Method is a combination of Method 3 and Method 1. For Tree A in Method 4, we fit a decision tree based on bagged training data with 25% positive rate. For Tree B in Method 4, we fit a decision tree based on training data with a 5% positive rate.



Method 4: 2-Step Sequential Tree consisting of Part A and Part B – One of 10 bagged models

Result 4: 2-Step Sequential Tree (Tree A - 25% Bagged, Tree B - 5% Simple Classification)

Method 4 performs extremely well. It achieves the same low misclassification rate of 0.1% as Method 1. Furthermore, since it is able to correctly classify 12.55% of patients in Tree A and thus save on the medical costs for these patients, it has a lower overall test cost of $47,907.02. This is a 14.46% reduction in the test cost, compared to Method 1. As such, this method the most ideal, capturing cost savings whilst ensuring the lowest possible misclassification rate of the models.

Method 4 Part A		Actual	
		Hypothyroid	Normal
Predicted	Hypothyroid	100	1649
	Normal	0	251

Method 4 Results (Part A): Classification on test data (2000 instances)

Method 4 Part B		Actual	
		Hypothyroid	Normal
Predicted	Hypothyroid	100	2
	Normal	0	1647

Method 4 Results (Part B): Classification on 'Hypothyroid' patients from Part A (1749 instances)

Method 4		Actual	
		Hypothyroid	Normal
Predicted	Hypothyroid	100	2
	Normal	0	1898

Method 4 Results: Final Classification on all test data (2000 instances)

Summary:

	Method 1	Method 2	Method 3	Method 4
Type	5% Positive Rate Simple Classification Tree	5% Positive Rate 2-Step Sequential Tree Tree A - 5% Simple Classification Tree Tree B - 5% Simple Classification Tree	25% Positive Rate 2-Step Sequential Tree (Bagged) 1. 25% Bagged Trees 2. 25% Bagged Trees	2-Step Sequential Tree 1. 25% Bagged Trees 2. 5% Simple Classification
Misclassification Rate	0.1%	4.35%	1.45%	0.1%
Medical Test Costs associated with classifying Test dataset	$55,058.58	$7,730.14	49,340.80 (Max)	$47,907.02

Comparison Table: Summary of our findings

From the comparison table, we see that Method 4 is unambiguously better than Method 1 and Method 2 as it achieves the same/lower misclassification rate but at lower costs. Method 4 is also better than Method 2, as it has a far lower misclassification rate. We view misclassification very harshly in our model, relative to the test cost saved. This reflects the severity of misdiagnosis, which threatens the patient's quality of life. As such, Method 4 is the most ideal, capturing cost savings whilst ensuring the lowest possible misclassification rate of the Methods.

Furthermore, to put our findings back into context, we note that status quo calls for all 5 medical tests to be run on all patients. For 2000 patients, this amounts to $202,780. Hence, comparing the test costs in Method 4 to the test costs in the status quo, we are able to decrease medical test costs by 76.37%. Method 4's misclassification rate of 0.1% likely compares favourably to the misdiagnosis rate of medical doctors.

4. CONCLUSION

This study has shown the potential of predictive analytics in improving patient outcomes and lowering costs. Using a two-step sequential tree, we are able to achieve a very low misclassification rate and lower tests costs. From the above analysis, we also learn that there are benefits and drawbacks to balancing an imbalanced training dataset. On one hand, balancing the training data prevents the information on 'positive' cases from being overwhelmed by that on the 'negative' cases. This allows the signal on 'positives' to be picked up by the decision tree, leading to be a better classification rate. On the other hand, when we train the decision tree on balanced data and solve for the optimum cut-off probability, this cut-off probability may be 'too low' for the test data, which has a lower positive rate. This results in the poor performance of the decision tree on test data. Thus, balancing an imbalanced training dataset should be done judiciously.

Through trial and error from Methods 1-3, we see that balancing the training dataset works well for Tree A in the 2-step sequential tree but not Tree B. This is because the increased signals from 'positive' cases is more important for Tree A, which is trained only on less informative observable data. Furthermore, since we prefer Tree A to be more cautious to avoid false-negatives, a low cut-off probability for diagnosing hypothyroidism is also desirable. Since Tree B is trained on the informative medical test data, it is able to classify the 'positive' cases accurately even from an imbalanced dataset. Hence, it is advantageous to train Tree B on the imbalanced training data. Thus, we learn that we must consider the nature of the problem and its requirements in deciding whether to balance an imbalanced training dataset.

However, our 2-step sequential tree is still not an optimised solution to the problem of test-cost classification as the decision tree's algorithm does not directly consider the medical tests costs. Moving forward, we suggest that further improvements can be found in the field of test-cost sensitive classification through algorithmic solutions (Yang et al, 2006). Test-cost sensitive decision tree packages can also be developed to create decisions trees which not only consider information gain (entropy decrease) in classification, but also the medical test costs of variables. This may result in optimized sequential trees that can minimise both misclassification and test costs.

5. BIBLIOGRAPHY

Source of data: Quinlan, J.R. (1993). UCI Machine Learning Repository [http://archive.ics.uci.edu/ml, http://archive.ics.uci.edu/ml/machine-learning-databases/thyroid-disease/ann-test.data, http://archive.ics.uci.edu/ml/machine-learning-databases/thyroid-disease/ann-train.data, http://archive.ics.uci.edu/ml/machine-learning-databases/thyroid-disease/costs/ann-thyroid.cost]. Irvine, CA: University of California, School of Information and Computer Science.

Test Cost: Accesalabs.com. (2019). Order Lab Tests Online - Walk In Blood Test Labs Near You | Accesa Labs ®. [online] Available at: https://www.accesalabs.com/

Cooper, D. S. (2001). Subclinical hypothyroidism. *New England Journal of Medicine*, *345*(4), 260-265.

Janke, A. T., Overbeek, D. L., Kocher, K. E., & Levy, P. D. (2016). Exploring the potential of predictive analytics and big data in emergency care. *Annals of emergency medicine*, *67*(2), 227-236.

Lavrač, N., Kononenko, I., Keravnou, E., Kukar, M., & Zupan, B. (1998). Intelligent data analysis for medical diagnosis: using machine learning and temporal abstraction. *AI Communications*, *11*(3, 4), 191-218.

Provost, F. and Fawcett, T. (2013). Data Science for Business. O'Reilly.

Raghupathi, W., & Raghupathi, V. (2013). An overview of health analytics. *J Health Med Informat*, *4*(132), 2.

Weetman, A. P. (1997). Fortnightly review: Hypothyroidism: screening and subclinical disease. *BMJ*, *314*(7088), 1175.

Wickham, H, and Grolemund, G. (2017). R for Data Science. O'Reilly.

Yang, Q., Ling, C., Chai, X., & Pan, R. (2006). Test-cost sensitive classification on data with missing values. *IEEE Transactions on Knowledge and Data Engineering*, *18*(5), 626-638.

6. APPENDIX

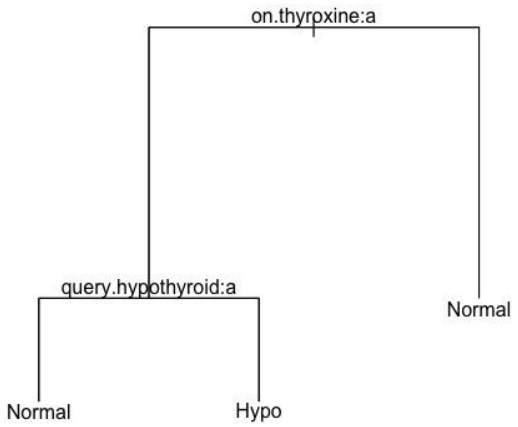
Dataset Attributes

Age	Numerical variable indicating the age of patients [0,1] Transformation by taking **Age***100
Sex	Female - indicated by '0' Male - indicated by '1'
On.thyroxine	Thyroxine refers to medication used to treat an underactive thyroid (Hypothyroidism) Negative - indicated by '0' Positive - indicated by '1'
Query on thyroxine	Insufficient information provided by the dataset
On antithyroid medication	Being on Antithyroid medication (also called “Thinoamides”) used to treat hyperthyroidism Negative - indicated by '0' Positive - indicated by '1'
Sick	Negative - indicated by '0' Positive - indicated by '1'
Pregnant	Pregnant status of a female Negative - indicated by '0' Positive - indicated by '1'
Thyroid Surgery	Prior history of having the thyroid gland removed either partially or entirely Negative - indicated by '0' Positive - indicated by '1'
I131.treatment	A form of treatment - Radioactive Iodine Therapy for Hyperthyroidism Have not undergone treatment - indicated by '0' Have undergone treatment - indicated by '1'
Query hypothyroidism	Insufficient information provided by the dataset
Query hyperthyroid	Insufficient information provided by the dataset
Lithium	A type of medication that is used for the treatment of mood disorders that can affect the thyroid function. Variable indicating use of medication Negative - indicated by '0' Positive - indicated by '1'
Goitre	Variable indicating the swelling of a patients’ neck resulting from the enlargement of the thyroid gland Negative - indicated by '0'

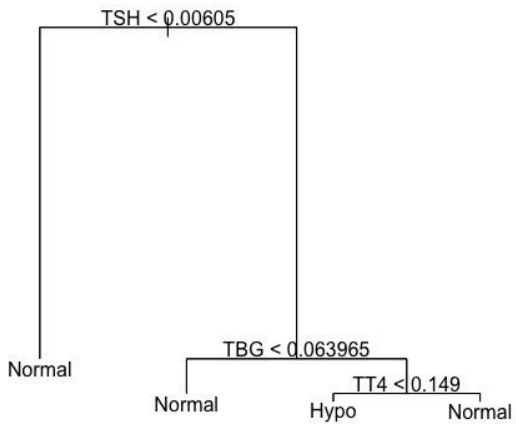
	Positive - indicated by '1'
Tumor	Variable indicating a tumor in the region of the pituitary Negative - indicated by '0' Positive - indicated by '1'
Hypopituitary	Variable indicating patient suffering from hypopituitary - where pituitary gland does not produce enough hormones Negative - indicated by '0' Positive - indicated by '1'
psych	Insufficient information provided by the dataset
TSH	Measure of Thyroid Stimulating Hormone
T3	Measure of a hormone called Triiodothyronine
TT4	Measure of Thyroxine
T4U	Measure of Thyroid Binding Capacity
TBG	Measure of Thyroxine Binding Globulin

Bagged Models for Method 3 (9 out of 10)

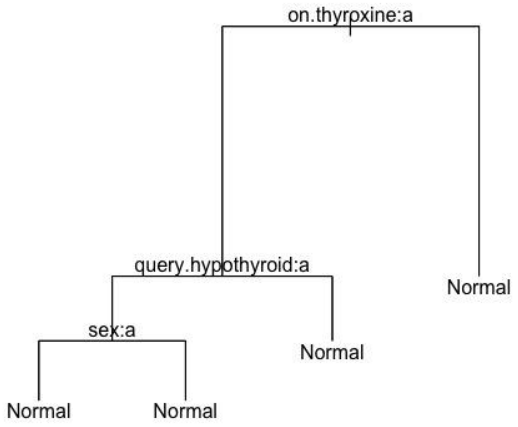
Tree 3A: Model 2



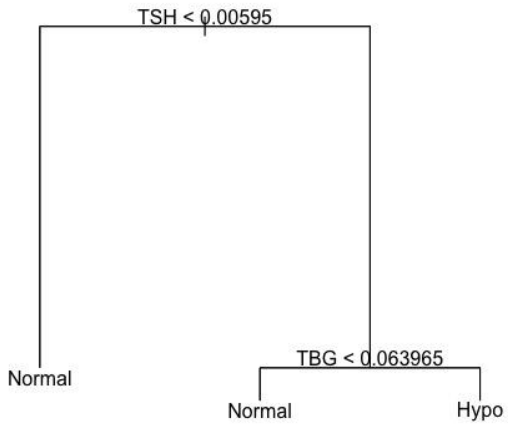
Tree 3B: Model 2



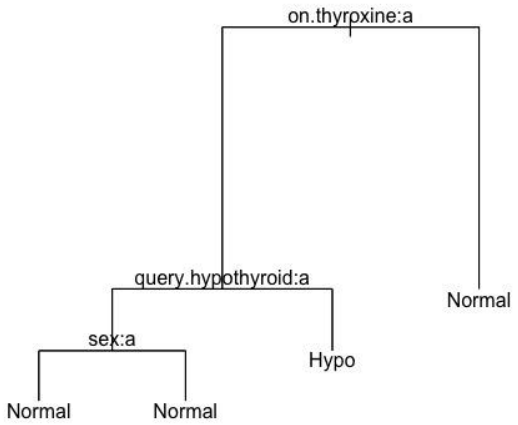
Tree 3A: Model 3



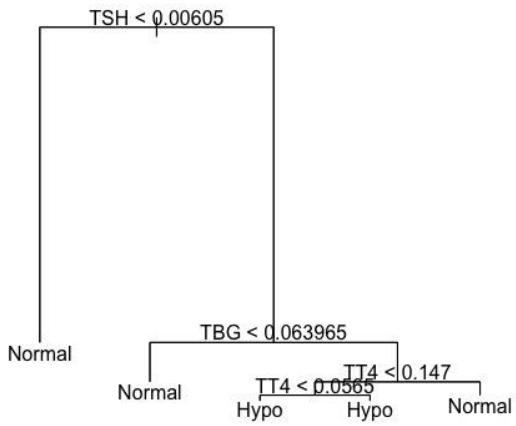
Tree 3B: Model 3



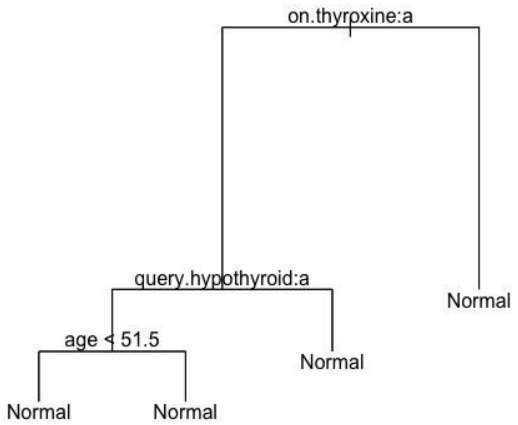
Tree 3A: Model 4



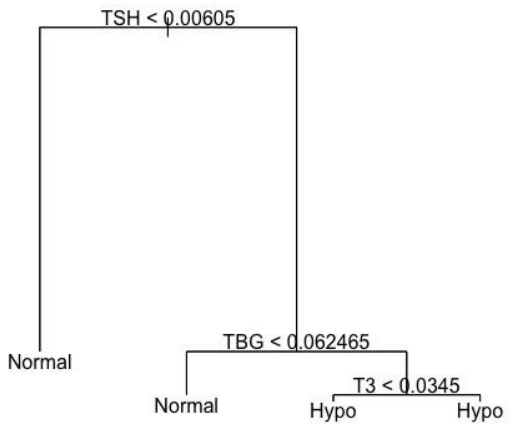
Tree 3B: Model 4



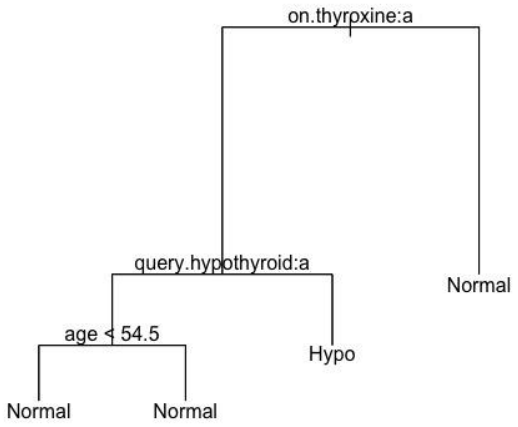
Tree 3A: Model 5



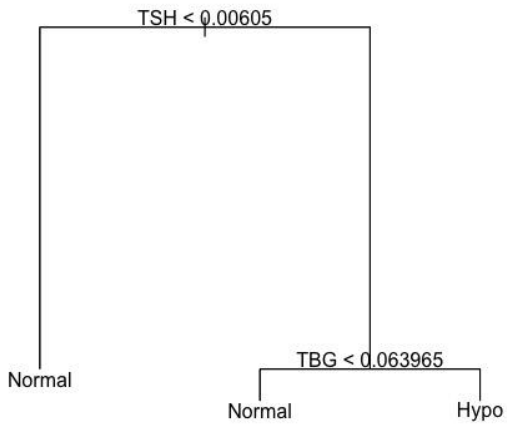
Tree 3B: Model 5



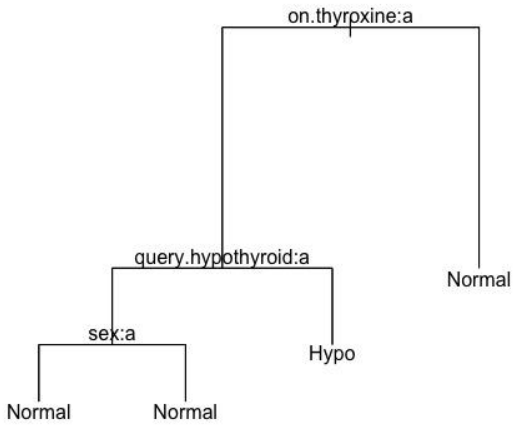
Tree 3A: Model 6



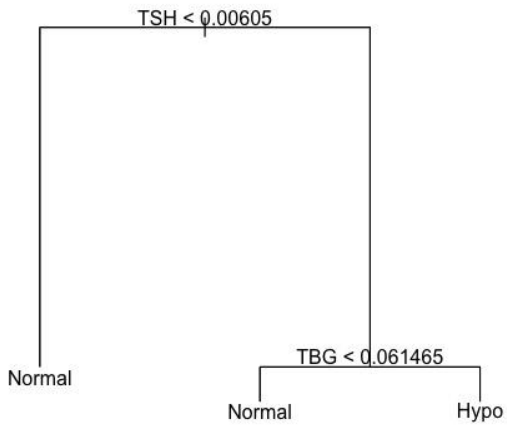
Tree 3B: Model 6



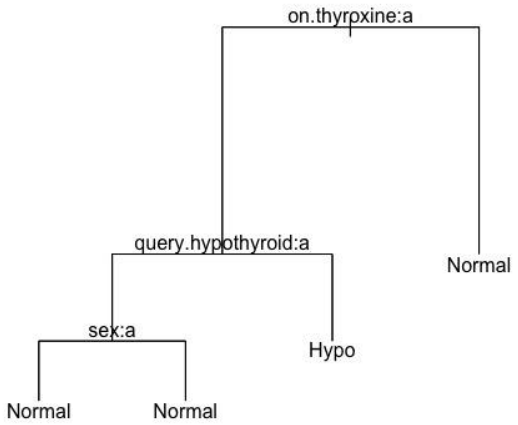
Tree 3A: Model 7



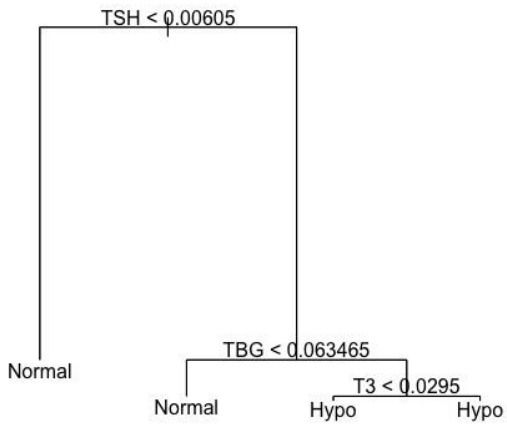
Tree 3B: Model 7



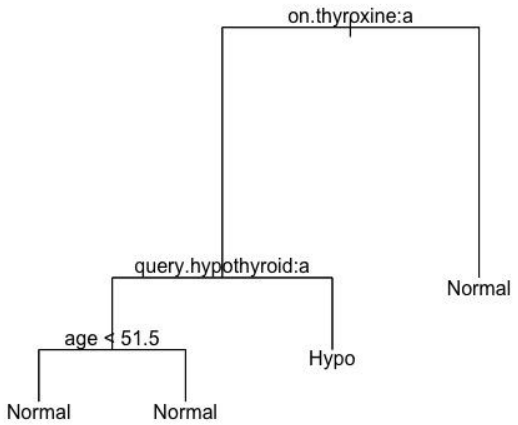
Tree 3A: Model 8



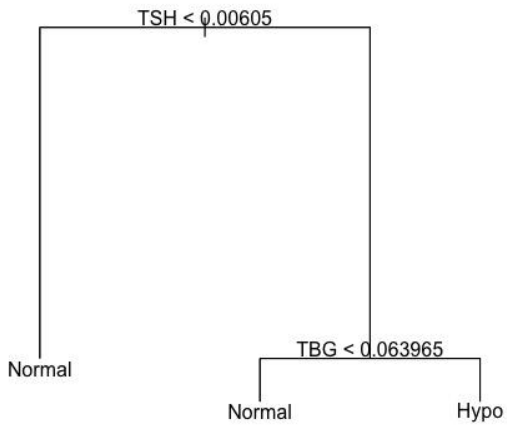
Tree 3B: Model 8



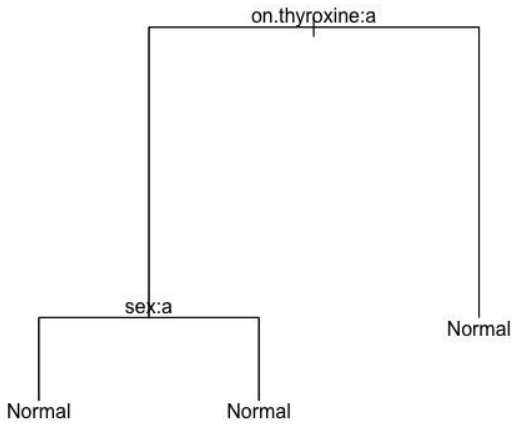
Tree 3A: Model 9



Tree 3B: Model 9



Tree 3A: Model 10



Tree 3B: Model 10