PES UNIVERSITY

DEPARTMENT OF ELECTRONICS AND COMMUNICATION ENGINEERING

A mini project report

TOPIC: ODI CRICKET PREDICTION

Submitted in Partial fulfilment of the Requirements for VI Semester
Bachelor of Technology in Electronics and Communication

Jan. - Apr. 2018

Under the guidance of

Dr. Koshy George K
Prof, Dept of ECE

TEAM MEMBERS

CHINMAY D KUCHINAD – 01FB16EEC081

CHIRANJEEVI N – 01FB16EEC083

DHEERAJ P ANIKAR – 01FB16EEC090

A machine learning approach to analyze ODI cricket predictors

Cricket has become one of the most popular sports in the world and craze for the game is massive in a country like India. In the recent years with improvement in technology the game in 21st century has become spectacular than ever before. The popularity of ODI cricket has been high for the past few decades. Broadcasters and cricket experts make best efforts to provide a great experience of the game to the viewers. Pre-match predictions keep the fans and public hooked to the game. In this project we are analyzing ODI cricket predictors using machine learning. For this the understanding of factors that affect the game is essential.

Factors affecting the prediction

By far the most crucial factor to be considered is the venue where the game is being played. Multiple studies have been carried out on home/away field advantages to the teams and have found considerable success in doing so. Also this approach has been taken further by incorporating the information about the continent to which the away team belongs.

But this may not always lead to desired results because the home factor may not be uniformly influential. Adopting this may lead to a high degree of over fitting, that too in sports prediction where the data available is not huge.

This can be overcome to a certain extent by taking into account additional factors, which may not necessarily be of uniform significance to the final outcome, but when considered together may lead to a closer approximation of actual results.

These additional factors affecting the predictions can broadly be classified as controllable, uncontrollable and partially controllable. While factors such as team combinations, on field strategies can be grouped under controllable, while venue , game type and coin toss fall under uncontrollable.

Coming to partially controllable factors, the choice of opting to bat or field can be considered one. Our approach considers these factors and a logistic regression model is to find the regression coefficients.

For our model we are considering home/away, toss result, choice after toss, game type as the parameters.

Data Preprocessing

We start out by considering the ESPN cricinfo dataset with information for matches between the years 1972 to 2015. Information about teams playing, toss outcomes, game type, choice after winning the toss is obtained from it.

Data cleaning is performed to remove any ambiguities if present. We have removed cricket matches that have been interrupted by rain, abandoned or ended in a tie as they affect our final outcome. The next step involves the team-wise partitioning of data. 9 such separate datasets are formed from the parent dataset.

This is followed by assigning discrete values of 0 or 1 to the above mentioned parameters as the 2 models discussed below. Once this is done the regression coefficients are calculated as described.

The logistic regression approach

Outcomes of bilateral games for a team ‘i’ can be modeled by the following logistic regression approach.

$$\ln\left(\frac{pi}{1-pi}\right) = \beta0 + \beta1 * HMi + \beta2 * DNi + \beta3 * TSi + \beta4 * BFi$$

where the predictors in this model are defined as follows.

HMi = 1 if it is a home game for the team i

0 otherwise

DNi = 1 if the game is a day game

0 if the game is a day-night game

TSi= 1 if the team i won the Toss

0 otherwise

BFi = 1 if the team i batted first

0 otherwise

Home field advantage includes cricket pitch, ground conditions and regional factors. Even though this model accounts for home-field advantage, it does not quantify the opposition’s foreign factor or the continent effect. The new variable we suggest here intends to account the home-field advantage of the home team with respect to the continent of the opposition team.

We have divided the cricket playing nations into their respective continents. Africa (South Africa, Kenya, and Zimbabwe), America (Canada and West Indies), Asia (Bangladesh, India, Pakistan, and Sri Lanka), Europe (England, Ireland, and Scotland) and Oceania (Australia and New Zealand).

The model 2 is the extension of model 1 to have five different continent parameters for each team including its own continent in place of HM.

Model 2 is as follows

$$\ln\left(\frac{pi}{1-pi}\right) = \beta 0 + \beta 1{*}DNi + \beta 2{*}TSi + \beta 3{*}BFi + \beta 4{*}AFRi + \beta 5{*}AMRi + \beta 6{*}ASAi + \beta 7{*}ERPi + \beta 8{*}OCNi$$

where continent variables for away teams are defined as

AFRi = 1 if the away team i∈ Africa

0 otherwise

AMRi = 1 if the away team i∈ America

0 otherwise

ASAi= 1 if the away team i∈ Asia

0 otherwise

ERPi = 1 if the away team i∈ Europe

0 otherwise

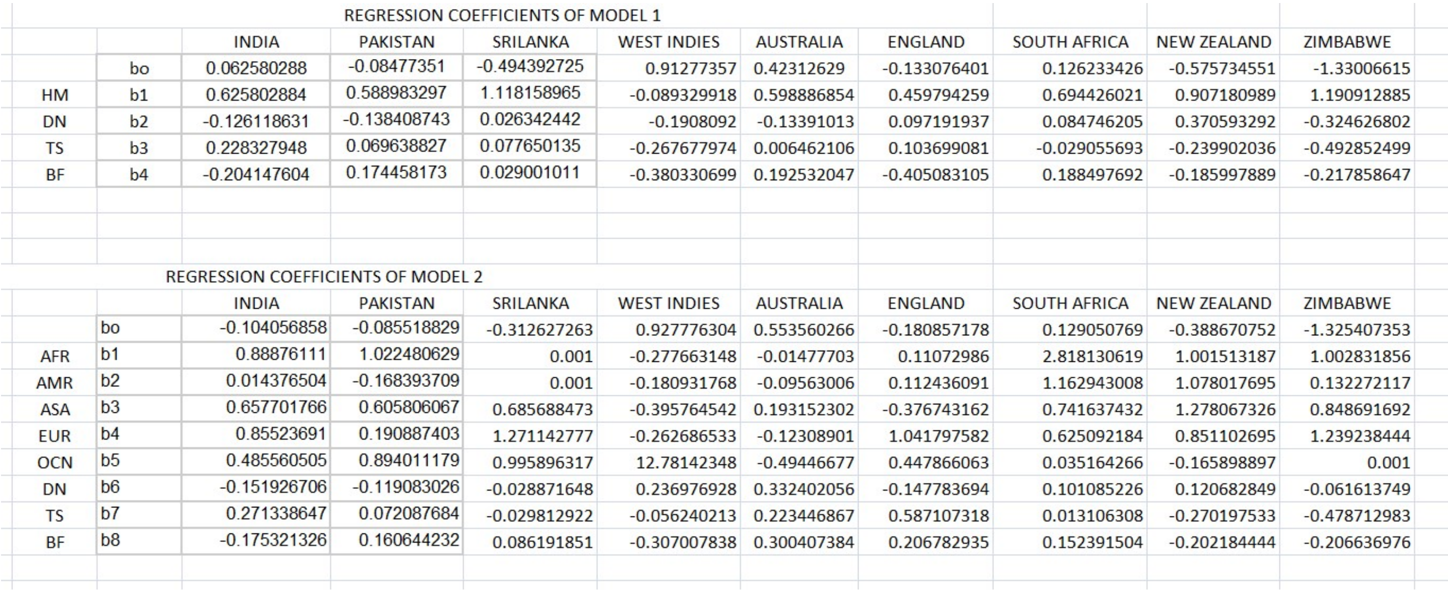
OCNi = 1 if the away team i∈ Oceania

0 otherwise

Finding the regression coefficients

We have found predictors using the following steps,

1. Find the logit (L) value using the above models
2. Find exponential of the logit and then calculate p(x) = $\frac{e^L}{1+e^L}$
3. Using the logistic regression formula find log likelihood of p(x)
   log likelihood = y*log(p(x)) + (1 – y)*(log(1 – p(x))
4. Using this, appropriate values of predictors/regression coefficients are calculated.



REGRESSION COEFFICIENTS OF MODEL 1

		INDIA	PAKISTAN	SRILANKA	WEST INDIES	AUSTRALIA	ENGLAND	SOUTH AFRICA	NEW ZEALAND	ZIMBABWE
	bo	0.062580288	-0.08477351	-0.494392725	0.91277357	0.42312629	-0.133076401	0.126233426	-0.575734551	-1.33006615
HM	b1	0.625802884	0.588983297	1.118158965	-0.089329918	0.598886854	0.459794259	0.694426021	0.907180989	1.190912885
DN	b2	-0.126118631	-0.138408743	0.026342442	-0.1908092	-0.13391013	0.097191937	0.084746205	0.370593292	-0.324626802
TS	b3	0.228327948	0.069638827	0.077650135	-0.267677974	0.006462106	0.103699081	-0.029055693	-0.239902036	-0.492852499
BF	b4	-0.204147604	0.174458173	0.029001011	-0.380330699	0.192532047	-0.405083105	0.188497692	-0.185997889	-0.217858647

REGRESSION COEFFICIENTS OF MODEL 2

		INDIA	PAKISTAN	SRILANKA	WEST INDIES	AUSTRALIA	ENGLAND	SOUTH AFRICA	NEW ZEALAND	ZIMBABWE
	bo	-0.104056858	-0.085518829	-0.312627263	0.927776304	0.553560266	-0.180857178	0.129050769	-0.388670752	-1.325407353
AFR	b1	0.88876111	1.022480629	0.001	-0.277663148	-0.01477703	0.11072986	2.818130619	1.001513187	1.002831856
AMR	b2	0.014376504	-0.168393709	0.001	-0.180931768	-0.09563006	0.112436091	1.162943008	1.078017695	0.132272117
ASA	b3	0.657701766	0.605806067	0.685688473	-0.395764542	0.193152302	-0.376743162	0.741637432	1.278067326	0.848691692
EUR	b4	0.85523691	0.190887403	1.271142777	-0.262686533	-0.12308901	1.041797582	0.625092184	0.851102695	1.239238444
OCN	b5	0.485560505	0.894011179	0.995896317	12.78142348	-0.49446677	0.447866063	0.035164266	-0.165898897	0.001
DN	b6	-0.151926706	-0.119083026	-0.028871648	0.236976928	0.332402056	-0.147783694	0.101085226	0.120682849	-0.061613749
TS	b7	0.271338647	0.072087684	-0.029812922	-0.056240213	0.223446867	0.587107318	0.013106308	-0.270197533	-0.478712983
BF	b8	-0.175321326	0.160644232	0.086191851	-0.307007838	0.300407384	0.206782935	0.152391504	-0.202184444	-0.206636976

Inferences that can be made from regression coefficients of model 1 and 2 are as follows

1. Home field advantage is significant for teams like India, New Zealand, Australia and South Africa.
2. Australia and South Africa win many games when they bat first
3. Winning toss plays major role in the wins of India and England
4. West Indies win good number of games when they bat first
5. India has good chance of winning against most of the teams that tour India
6. Australia receive a good fight from England and South Africa while playing at home
7. South Africa wins many games against all the Asian and European nations

These are some of the many inferences that can be made.

In the next section we look at the CART – Classification and Regression Trees approach

CART – CLASSIFICATION AND REGRESSION TREES

CART is a popular ML technique that can be to predict outcomes and also to acknowledge the importance of predictors. The CART is a binary decision tree algorithm which recursively partitions data. Importantly, it can handle both continuous and categorical responses and predictors. In general, the trees are grown to the maximum possible size and then get rid of unimportant predictors (pruning) sequentially depending on a cost-complexity measure. Usually, this algorithm produces a set of pruned trees and finally selects the final tree as the best predictive model.

Decision Tree Algorithm

Decision Tree Classifier is a simple and widely used classification technique. It applies a straightforward idea to solve the classification problem. Decision Tree Classifier poses a series of carefully crafted questions about the attributes of the test record. Each time it receives an answer, a follow-up question is asked until a conclusion about the class label of the record is reached.

Gini impurity

The order or the relative importance of these questions are kept track of using certain parameter which in our case is gini impurity. Gini impurity is a measure of how often a randomly chosen element from the set would be incorrectly labeled if it was randomly labeled according to the distribution of labels in the subset.

Gini impurity function,

$$\mathbf{i(t) = 1 - p(t)^2 - (1 - p(t))^2}$$

where p(t) is the conditional probability of class 1 provided the current node is t. The change or the gain of the impurity function due to a split of the parent node (tp) into left and right children nodes t_L and t_R,

respectively, is

$$\mathbf{\Delta\ i(t) = i(tp) - qi(t_L) - (1 - q)i(t_R)}$$

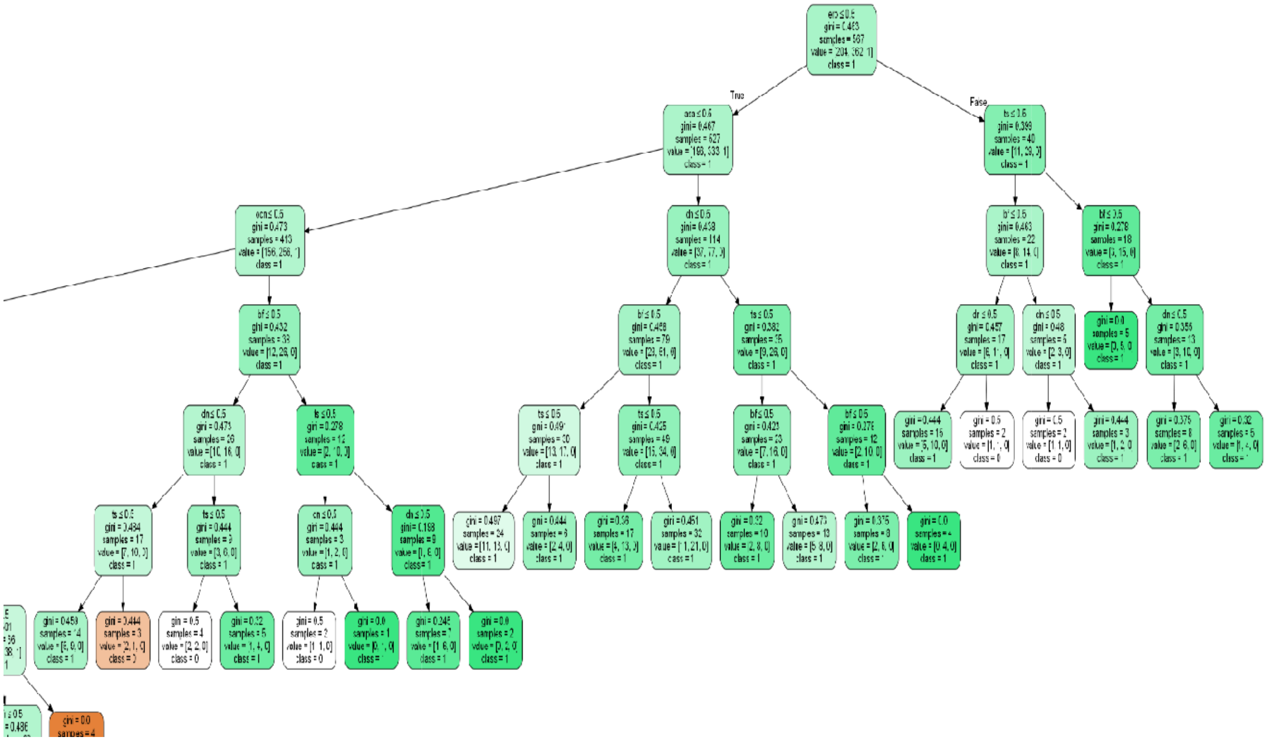
where q is the probability of the instances going to the left node. CART finds the best splitting criteria which maximizes the gain in impurity measure Δi(t).

Implementing the decision tree

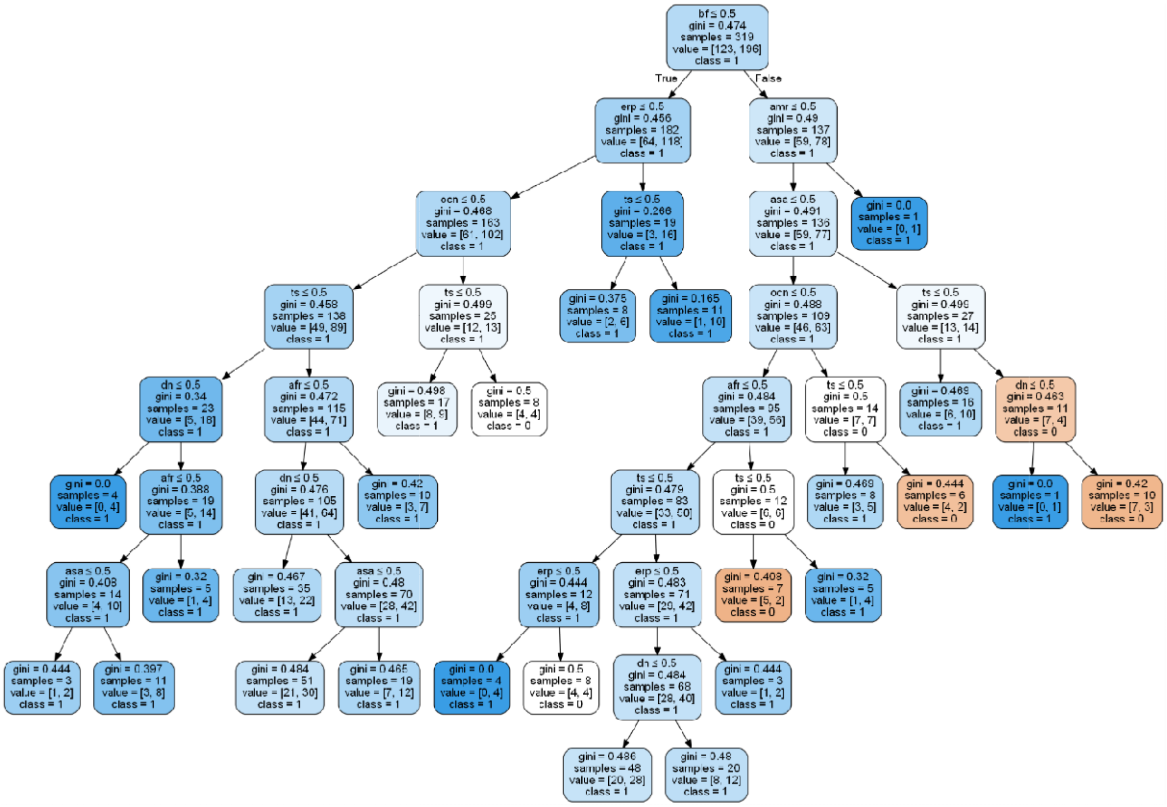
We use two approaches. One using the predefined function in scikit learn and the other involving writing the function on our own.

The scikit learn tree by default branches to the left child node for true cases and right child node for false cases. The parameters are placed hierarchically based on the gini impurity values and the relative importance of the parameters with respect to the final outcome is determined.

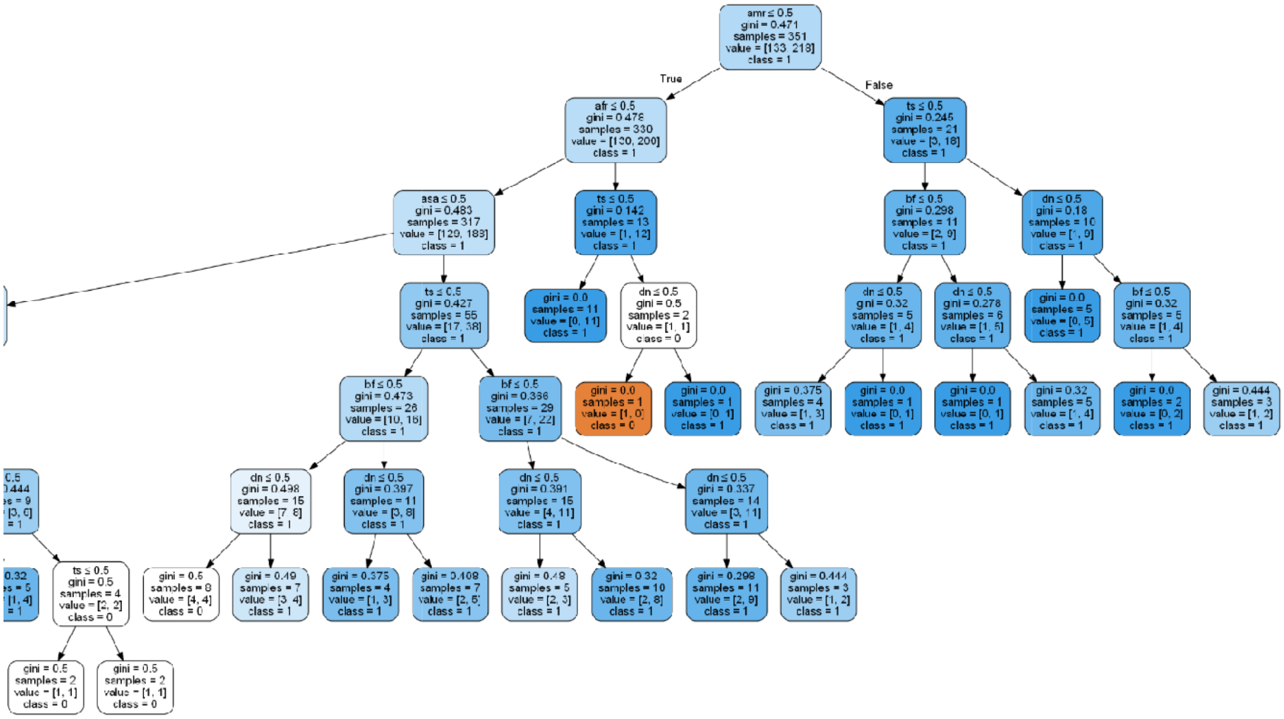
Australia


erp ≤ 0.5 gini = 0.463 samples = 567 value = [204, 362, 1] class = 1
True
False
asa ≤ 0.5 gini = 0.467 samples = 527 value = [193, 333, 1] class = 1
ts ≤ 0.5 gini = 0.399 samples = 40 value = [11, 29, 0] class = 1
ocn ≤ 0.5 gini = 0.473 samples = 413 value = [156, 256, 1] class = 1
dn ≤ 0.5 gini = 0.438 samples = 114 value = [37, 77, 0] class = 1
bf ≤ 0.5 gini = 0.463 samples = 22 value = [8, 14, 0] class = 1
bf ≤ 0.5 gini = 0.278 samples = 18 value = [3, 15, 0] class = 1
bf ≤ 0.5 gini = 0.432 samples = 38 value = [12, 26, 0] class = 1
bf ≤ 0.5 gini = 0.468 samples = 79 value = [28, 51, 0] class = 1
ts ≤ 0.5 gini = 0.382 samples = 35 value = [9, 26, 0] class = 1
dr ≤ 0.5 gini = 0.457 samples = 17 value = [6, 11, 0] class = 1
dn ≤ 0.5 gini = 0.48 samples = 5 value = [2, 3, 0] class = 1
gini = 0.0 samples = 5 value = [0, 5, 0] class = 1
dn ≤ 0.5 gini = 0.355 samples = 13 value = [3, 10, 0] class = 1
dn ≤ 0.5 gini = 0.473 samples = 26 value = [10, 16, 0] class = 1
ts ≤ 0.5 gini = 0.278 samples = 12 value = [2, 10, 0] class = 1
ts ≤ 0.5 gini = 0.491 samples = 30 value = [13, 17, 0] class = 1
ts ≤ 0.5 gini = 0.425 samples = 49 value = [15, 34, 0] class = 1
bf ≤ 0.5 gini = 0.423 samples = 23 value = [7, 16, 0] class = 1
bf ≤ 0.5 gini = 0.278 samples = 12 value = [2, 10, 0] class = 1
gini = 0.444 samples = 15 value = [5, 10, 0] class = 1
gini = 0.5 samples = 2 value = [1, 1, 0] class = 0
gini = 0.5 samples = 2 value = [1, 1, 0] class = 0
gini = 0.444 samples = 3 value = [1, 2, 0] class = 1
gini = 0.375 samples = 8 value = [2, 6, 0] class = 1
gini = 0.32 samples = 5 value = [1, 4, 0] class = 1
ts ≤ 0.5 gini = 0.484 samples = 17 value = [7, 10, 0] class = 1
ts ≤ 0.5 gini = 0.444 samples = 9 value = [3, 6, 0] class = 1
cn ≤ 0.5 gini = 0.444 samples = 3 value = [1, 2, 0] class = 1
dn ≤ 0.5 gini = 0.198 samples = 9 value = [1, 8, 0] class = 1
gini = 0.497 samples = 24 value = [11, 13, 0] class = 1
gini = 0.444 samples = 6 value = [2, 4, 0] class = 1
gini = 0.36 samples = 17 value = [4, 13, 0] class = 1
gini = 0.451 samples = 32 value = [11, 21, 0] class = 1
gini = 0.32 samples = 10 value = [2, 8, 0] class = 1
gini = 0.473 samples = 13 value = [5, 8, 0] class = 1
gini = 0.375 samples = 8 value = [2, 6, 0] class = 1
gini = 0.0 samples = 4 value = [0, 4, 0] class = 1
gini = 0.459 samples = 14 value = [5, 9, 0] class = 1
gini = 0.444 samples = 3 value = [2, 1, 0] class = 0
gini = 0.5 samples = 4 value = [2, 2, 0] class = 0
gini = 0.32 samples = 5 value = [1, 4, 0] class = 1
gini = 0.5 samples = 2 value = [1, 1, 0] class = 0
gini = 0.0 samples = 1 value = [0, 1, 0] class = 1
gini = 0.245 samples = 7 value = [1, 6, 0] class = 1
gini = 0.0 samples = 2 value = [0, 2, 0] class = 1
gini = 0.0 samples = 4

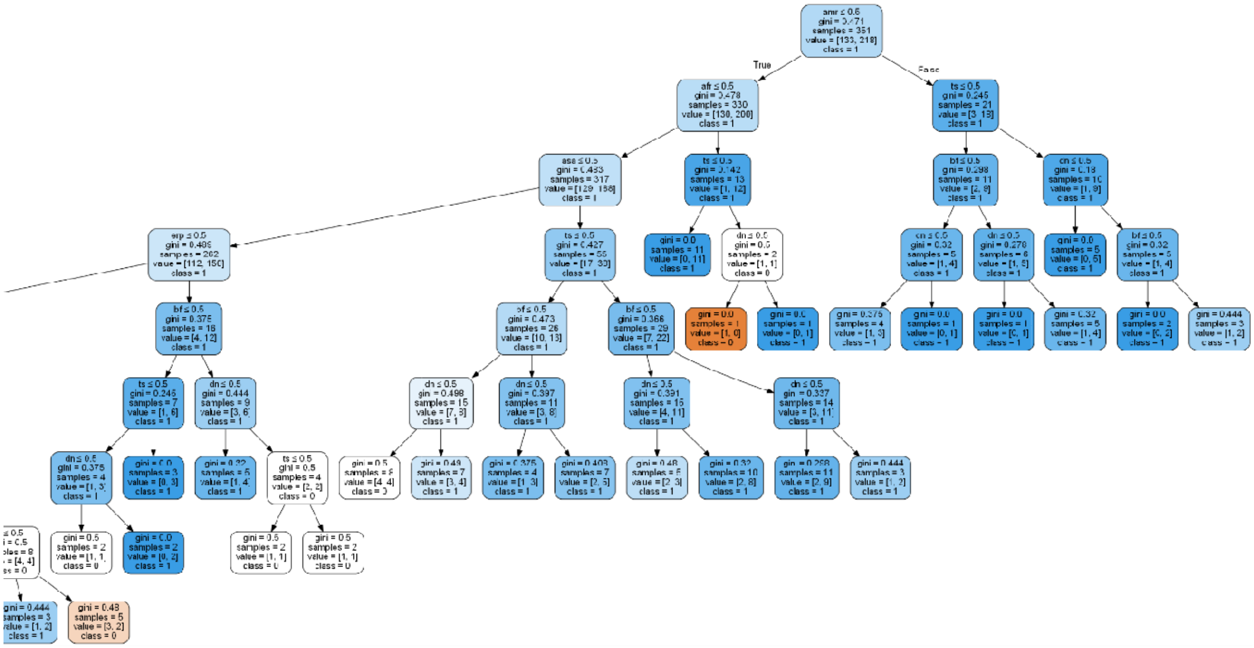
West Indies


bf ≤ 0.5 gini = 0.474 samples = 319 value = [123, 196] class = 1
True
False
erp ≤ 0.5 gini = 0.456 samples = 182 value = [64, 118] class = 1
amr ≤ 0.5 gini = 0.49 samples = 137 value = [59, 78] class = 1
ocn ≤ 0.5 gini = 0.468 samples = 163 value = [61, 102] class = 1
ts ≤ 0.5 gini = 0.266 samples = 19 value = [3, 16] class = 1
asa ≤ 0.5 gini = 0.491 samples = 136 value = [59, 77] class = 1
gini = 0.0 samples = 1 value = [0, 1] class = 1
ts ≤ 0.5 gini = 0.458 samples = 138 value = [49, 89] class = 1
ts ≤ 0.5 gini = 0.499 samples = 25 value = [12, 13] class = 1
gini = 0.375 samples = 8 value = [2, 6] class = 1
gini = 0.165 samples = 11 value = [1, 10] class = 1
ocn ≤ 0.5 gini = 0.488 samples = 109 value = [46, 63] class = 1
ts ≤ 0.5 gini = 0.499 samples = 27 value = [13, 14] class = 1
dn ≤ 0.5 gini = 0.34 samples = 23 value = [5, 18] class = 1
afr ≤ 0.5 gini = 0.472 samples = 115 value = [44, 71] class = 1
gini = 0.498 samples = 17 value = [8, 9] class = 1
gini = 0.5 samples = 8 value = [4, 4] class = 0
afr ≤ 0.5 gini = 0.484 samples = 95 value = [39, 56] class = 1
ts ≤ 0.5 gini = 0.5 samples = 14 value = [7, 7] class = 0
gini = 0.469 samples = 16 value = [6, 10] class = 1
dn ≤ 0.5 gini = 0.463 samples = 11 value = [7, 4] class = 0
gini = 0.0 samples = 4 value = [0, 4] class = 1
afr ≤ 0.5 gini = 0.388 samples = 19 value = [5, 14] class = 1
dn ≤ 0.5 gini = 0.476 samples = 105 value = [41, 64] class = 1
gini = 0.42 samples = 10 value = [3, 7] class = 1
ts ≤ 0.5 gini = 0.479 samples = 83 value = [33, 50] class = 1
ts ≤ 0.5 gini = 0.5 samples = 12 value = [6, 6] class = 0
gini = 0.469 samples = 8 value = [3, 5] class = 1
gini = 0.444 samples = 6 value = [4, 2] class = 0
gini = 0.0 samples = 1 value = [0, 1] class = 1
gini = 0.42 samples = 10 value = [7, 3] class = 0
asa ≤ 0.5 gini = 0.408 samples = 14 value = [4, 10] class = 1
gini = 0.32 samples = 5 value = [1, 4] class = 1
gini = 0.467 samples = 35 value = [13, 22] class = 1
asa ≤ 0.5 gini = 0.48 samples = 70 value = [28, 42] class = 1
erp ≤ 0.5 gini = 0.444 samples = 12 value = [4, 8] class = 1
erp ≤ 0.5 gini = 0.483 samples = 71 value = [29, 42] class = 1
gini = 0.408 samples = 7 value = [5, 2] class = 0
gini = 0.32 samples = 5 value = [1, 4] class = 1
gini = 0.444 samples = 3 value = [1, 2] class = 1
gini = 0.397 samples = 11 value = [3, 8] class = 1
gini = 0.484 samples = 51 value = [21, 30] class = 1
gini = 0.465 samples = 19 value = [7, 12] class = 1
gini = 0.0 samples = 4 value = [0, 4] class = 1
gini = 0.5 samples = 8 value = [4, 4] class = 0
dn ≤ 0.5 gini = 0.484 samples = 68 value = [28, 40] class = 1
gini = 0.444 samples = 3 value = [1, 2] class = 1
gini = 0.486 samples = 48 value = [20, 28] class = 1
gini = 0.48 samples = 20 value = [8, 12] class = 1

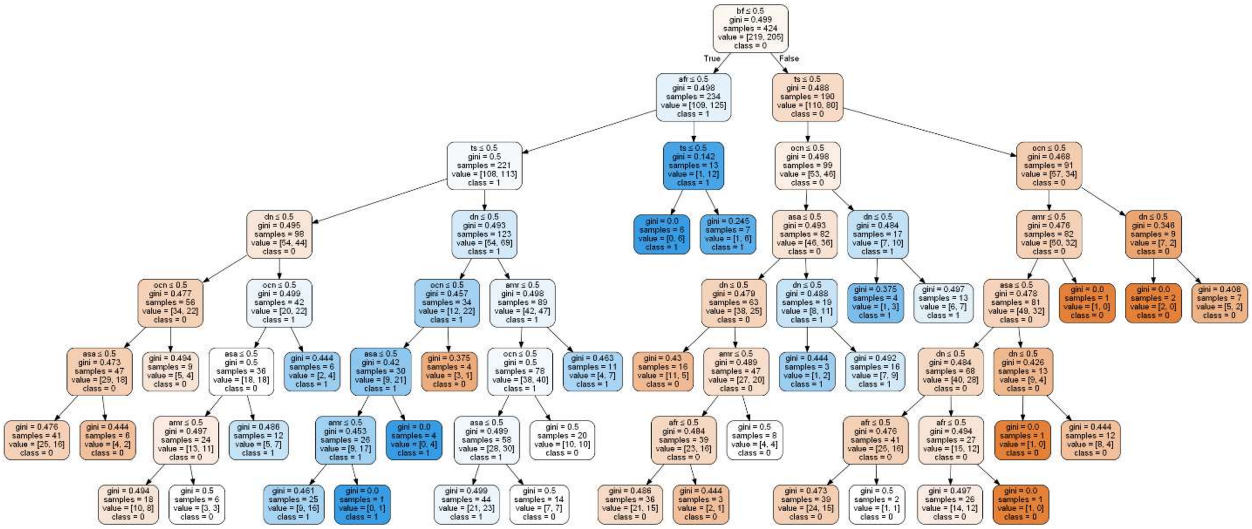
South Africa


amr ≤ 0.5 gini = 0.471 samples = 351 value = [133, 218] class = 1
True
False
afr ≤ 0.5 gini = 0.478 samples = 330 value = [130, 200] class = 1
ts ≤ 0.5 gini = 0.245 samples = 21 value = [3, 18] class = 1
asa ≤ 0.5 gini = 0.483 samples = 317 value = [129, 188] class = 1
ts ≤ 0.5 gini = 0.142 samples = 13 value = [1, 12] class = 1
bf ≤ 0.5 gini = 0.298 samples = 11 value = [2, 9] class = 1
dn ≤ 0.5 gini = 0.18 samples = 10 value = [1, 9] class = 1
ts ≤ 0.5 gini = 0.427 samples = 55 value = [17, 38] class = 1
gini = 0.0 samples = 11 value = [0, 11] class = 1
dn ≤ 0.5 gini = 0.5 samples = 2 value = [1, 1] class = 0
dn ≤ 0.5 gini = 0.32 samples = 5 value = [1, 4] class = 1
dn ≤ 0.5 gini = 0.278 samples = 6 value = [1, 5] class = 1
gini = 0.0 samples = 5 value = [0, 5] class = 1
bf ≤ 0.5 gini = 0.32 samples = 5 value = [1, 4] class = 1
bf ≤ 0.5 gini = 0.473 samples = 26 value = [10, 16] class = 1
bf ≤ 0.5 gini = 0.366 samples = 29 value = [7, 22] class = 1
gini = 0.0 samples = 1 value = [1, 0] class = 0
gini = 0.0 samples = 1 value = [0, 1] class = 1
gini = 0.375 samples = 4 value = [1, 3] class = 1
gini = 0.0 samples = 1 value = [0, 1] class = 1
gini = 0.0 samples = 1 value = [0, 1] class = 1
gini = 0.32 samples = 5 value = [1, 4] class = 1
gini = 0.0 samples = 2 value = [0, 2] class = 1
gini = 0.444 samples = 3 value = [1, 2] class = 1
0.5 0.444 s = 9 [3, 6] = 1
dn ≤ 0.5 gini = 0.498 samples = 15 value = [7, 8] class = 1
dn ≤ 0.5 gini = 0.397 samples = 11 value = [3, 8] class = 1
dn ≤ 0.5 gini = 0.391 samples = 15 value = [4, 11] class = 1
dn ≤ 0.5 gini = 0.337 samples = 14 value = [3, 11] class = 1
0.32 s = 5 [1, 4] = 1
ts ≤ 0.5 gini = 0.5 samples = 4 value = [2, 2] class = 0
gini = 0.5 samples = 8 value = [4, 4] class = 0
gini = 0.49 samples = 7 value = [3, 4] class = 1
gini = 0.375 samples = 4 value = [1, 3] class = 1
gini = 0.408 samples = 7 value = [2, 5] class = 1
gini = 0.48 samples = 5 value = [2, 3] class = 1
gini = 0.32 samples = 10 value = [2, 8] class = 1
gini = 0.298 samples = 11 value = [2, 9] class = 1
gini = 0.444 samples = 3 value = [1, 2] class = 1
gini = 0.5 samples = 2 value = [1, 1] class = 0
gini = 0.5 samples = 2 value = [1, 1] class = 0

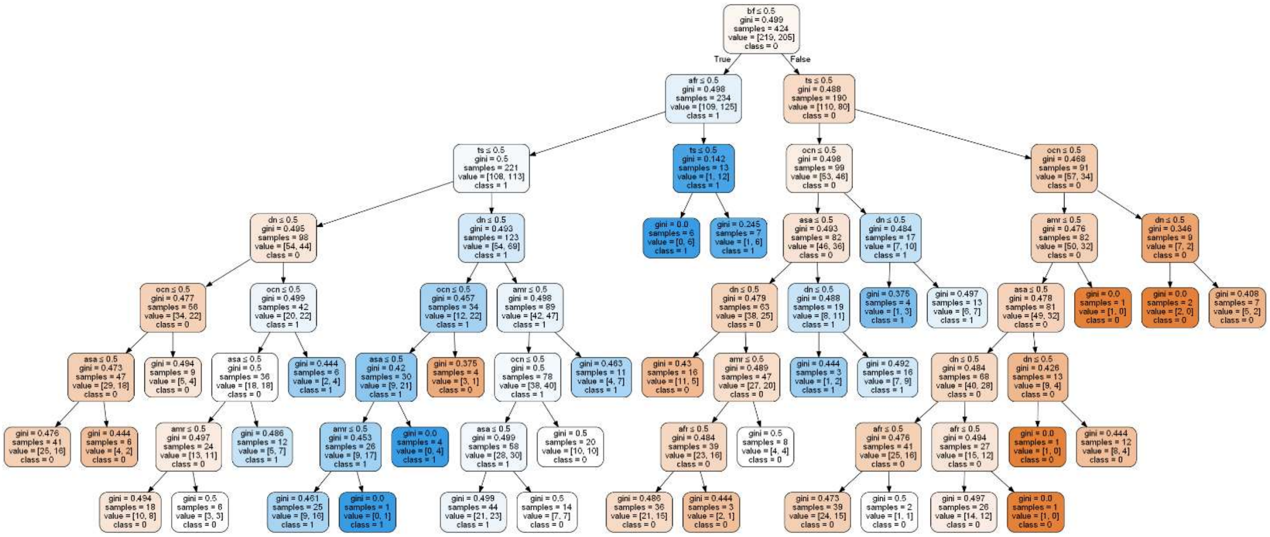
New Zealand


amr ≤ 0.5 gini = 0.471 samples = 351 value = [133, 218] class = 1
True
False
afr ≤ 0.5 gini = 0.478 samples = 330 value = [130, 200] class = 1
ts ≤ 0.5 gini = 0.245 samples = 21 value = [3, 18] class = 1
asa ≤ 0.5 gini = 0.483 samples = 317 value = [129, 188] class = 1
ts ≤ 0.5 gini = 0.142 samples = 13 value = [1, 12] class = 1
bf ≤ 0.5 gini = 0.298 samples = 11 value = [2, 9] class = 1
dn ≤ 0.5 gini = 0.18 samples = 10 value = [1, 9] class = 1
erp ≤ 0.5 gini = 0.489 samples = 262 value = [112, 150] class = 1
ts ≤ 0.5 gini = 0.427 samples = 55 value = [17, 38] class = 1
gini = 0.0 samples = 11 value = [0, 11] class = 1
dn ≤ 0.5 gini = 0.5 samples = 2 value = [1, 1] class = 0
dn ≤ 0.5 gini = 0.32 samples = 5 value = [1, 4] class = 1
dn ≤ 0.5 gini = 0.278 samples = 6 value = [1, 5] class = 1
gini = 0.0 samples = 5 value = [0, 5] class = 1
bf ≤ 0.5 gini = 0.32 samples = 5 value = [1, 4] class = 1
bf ≤ 0.5 gini = 0.375 samples = 16 value = [4, 12] class = 1
bf ≤ 0.5 gini = 0.473 samples = 26 value = [10, 16] class = 1
bf ≤ 0.5 gini = 0.366 samples = 29 value = [7, 22] class = 1
gini = 0.0 samples = 1 value = [1, 0] class = 0
gini = 0.0 samples = 1 value = [0, 1] class = 1
gini = 0.375 samples = 4 value = [1, 3] class = 1
gini = 0.0 samples = 1 value = [0, 1] class = 1
gini = 0.0 samples = 1 value = [0, 1] class = 1
gini = 0.32 samples = 5 value = [1, 4] class = 1
gini = 0.0 samples = 2 value = [0, 2] class = 1
gini = 0.444 samples = 3 value = [1, 2] class = 1
ts ≤ 0.5 gini = 0.245 samples = 7 value = [1, 6] class = 1
dn ≤ 0.5 gini = 0.444 samples = 9 value = [3, 6] class = 1
dn ≤ 0.5 gini = 0.498 samples = 15 value = [7, 8] class = 1
dn ≤ 0.5 gini = 0.397 samples = 11 value = [3, 8] class = 1
dn ≤ 0.5 gini = 0.391 samples = 15 value = [4, 11] class = 1
dn ≤ 0.5 gini = 0.337 samples = 14 value = [3, 11] class = 1
dn ≤ 0.5 gini = 0.375 samples = 4 value = [1, 3] class = 1
gini = 0.0 samples = 3 value = [0, 3] class = 1
gini = 0.32 samples = 5 value = [1, 4] class = 1
ts ≤ 0.5 gini = 0.5 samples = 4 value = [2, 2] class = 0
gini = 0.5 samples = 8 value = [4, 4] class = 0
gini = 0.49 samples = 7 value = [3, 4] class = 1
gini = 0.375 samples = 4 value = [1, 3] class = 1
gini = 0.408 samples = 7 value = [2, 5] class = 1
gini = 0.48 samples = 5 value = [2, 3] class = 1
gini = 0.32 samples = 10 value = [2, 8] class = 1
gini = 0.298 samples = 11 value = [2, 9] class = 1
gini = 0.444 samples = 3 value = [1, 2] class = 1
≤ 0.5 i = 0.5 ples = 8 = [4, 4] ss = 0
gini = 0.5 samples = 2 value = [1, 1] class = 0
gini = 0.0 samples = 2 value = [0, 2] class = 1
gini = 0.5 samples = 2 value = [1, 1] class = 0
gini = 0.5 samples = 2 value = [1, 1] class = 0
gini = 0.444 samples = 3 value = [1, 2] class = 1
gini = 0.48 samples = 5 value = [3, 2] class = 0

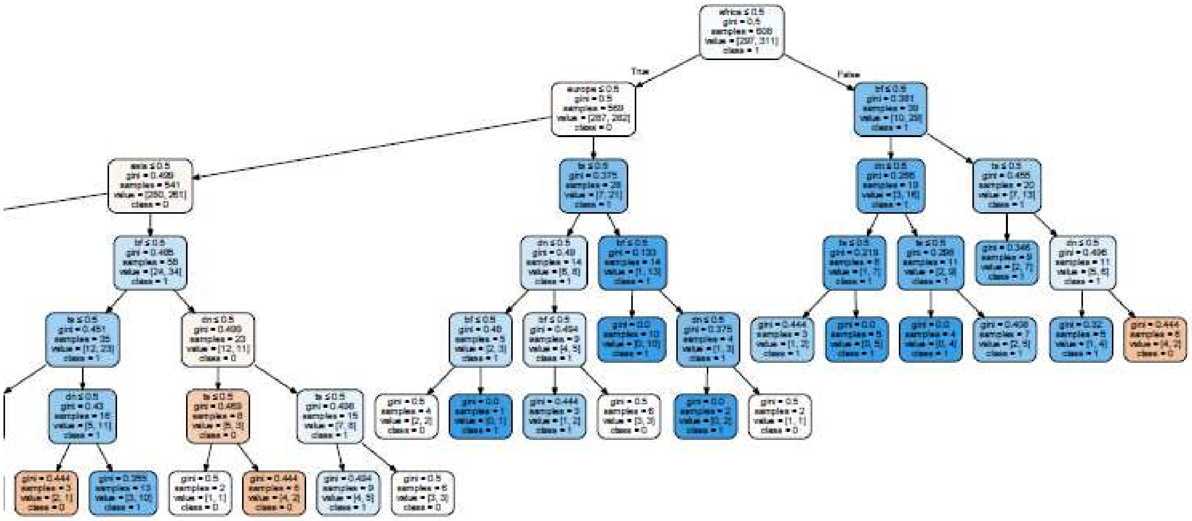
England


bf ≤ 0.5 gini = 0.499 samples = 424 value = [219, 205] class = 0
True
False
afr ≤ 0.5 gini = 0.498 samples = 234 value = [109, 125] class = 1
ts ≤ 0.5 gini = 0.488 samples = 190 value = [110, 80] class = 0
ts ≤ 0.5 gini = 0.5 samples = 221 value = [108, 113] class = 1
ts ≤ 0.5 gini = 0.142 samples = 13 value = [1, 12] class = 1
ocn ≤ 0.5 gini = 0.498 samples = 99 value = [53, 46] class = 0
ocn ≤ 0.5 gini = 0.468 samples = 91 value = [57, 34] class = 0
dn ≤ 0.5 gini = 0.495 samples = 98 value = [54, 44] class = 0
dn ≤ 0.5 gini = 0.493 samples = 123 value = [54, 69] class = 1
gini = 0.0 samples = 6 value = [0, 6] class = 1
gini = 0.245 samples = 7 value = [1, 6] class = 1
asa ≤ 0.5 gini = 0.493 samples = 82 value = [46, 36] class = 0
dn ≤ 0.5 gini = 0.484 samples = 17 value = [7, 10] class = 1
amr ≤ 0.5 gini = 0.476 samples = 82 value = [50, 32] class = 0
dn ≤ 0.5 gini = 0.346 samples = 9 value = [7, 2] class = 0
ocn ≤ 0.5 gini = 0.477 samples = 56 value = [34, 22] class = 0
ocn ≤ 0.5 gini = 0.499 samples = 42 value = [20, 22] class = 1
ocn ≤ 0.5 gini = 0.457 samples = 34 value = [12, 22] class = 1
amr ≤ 0.5 gini = 0.498 samples = 89 value = [42, 47] class = 1
dn ≤ 0.5 gini = 0.479 samples = 63 value = [38, 25] class = 0
dn ≤ 0.5 gini = 0.488 samples = 19 value = [8, 11] class = 1
gini = 0.375 samples = 4 value = [1, 3] class = 1
gini = 0.497 samples = 13 value = [6, 7] class = 1
asa ≤ 0.5 gini = 0.478 samples = 81 value = [49, 32] class = 0
gini = 0.0 samples = 1 value = [1, 0] class = 0
gini = 0.0 samples = 2 value = [2, 0] class = 0
gini = 0.408 samples = 7 value = [5, 2] class = 0
asa ≤ 0.5 gini = 0.473 samples = 47 value = [29, 18] class = 0
gini = 0.494 samples = 9 value = [5, 4] class = 0
asa ≤ 0.5 gini = 0.5 samples = 36 value = [18, 18] class = 0
gini = 0.444 samples = 6 value = [2, 4] class = 1
asa ≤ 0.5 gini = 0.42 samples = 30 value = [9, 21] class = 1
gini = 0.375 samples = 4 value = [3, 1] class = 0
ocn ≤ 0.5 gini = 0.5 samples = 78 value = [38, 40] class = 1
gini = 0.463 samples = 11 value = [4, 7] class = 1
gini = 0.43 samples = 16 value = [11, 5] class = 0
amr ≤ 0.5 gini = 0.489 samples = 47 value = [27, 20] class = 0
gini = 0.444 samples = 3 value = [1, 2] class = 1
gini = 0.492 samples = 16 value = [7, 9] class = 1
dn ≤ 0.5 gini = 0.484 samples = 68 value = [40, 28] class = 0
dn ≤ 0.5 gini = 0.426 samples = 13 value = [9, 4] class = 0
gini = 0.476 samples = 41 value = [25, 16] class = 0
gini = 0.444 samples = 6 value = [4, 2] class = 0
amr ≤ 0.5 gini = 0.497 samples = 24 value = [13, 11] class = 0
gini = 0.486 samples = 12 value = [5, 7] class = 1
amr ≤ 0.5 gini = 0.453 samples = 26 value = [9, 17] class = 1
gini = 0.0 samples = 4 value = [0, 4] class = 1
asa ≤ 0.5 gini = 0.499 samples = 58 value = [28, 30] class = 1
gini = 0.5 samples = 20 value = [10, 10] class = 0
afr ≤ 0.5 gini = 0.484 samples = 39 value = [23, 16] class = 0
gini = 0.5 samples = 8 value = [4, 4] class = 0
afr ≤ 0.5 gini = 0.476 samples = 41 value = [25, 16] class = 0
afr ≤ 0.5 gini = 0.494 samples = 27 value = [15, 12] class = 0
gini = 0.0 samples = 1 value = [1, 0] class = 0
gini = 0.444 samples = 12 value = [8, 4] class = 0
gini = 0.494 samples = 18 value = [10, 8] class = 0
gini = 0.5 samples = 6 value = [3, 3] class = 0
gini = 0.461 samples = 25 value = [9, 16] class = 1
gini = 0.0 samples = 1 value = [0, 1] class = 1
gini = 0.499 samples = 44 value = [21, 23] class = 1
gini = 0.5 samples = 14 value = [7, 7] class = 0
gini = 0.486 samples = 36 value = [21, 15] class = 0
gini = 0.444 samples = 3 value = [2, 1] class = 0
gini = 0.473 samples = 39 value = [24, 15] class = 0
gini = 0.5 samples = 2 value = [1, 1] class = 0
gini = 0.497 samples = 26 value = [14, 12] class = 0
gini = 0.0 samples = 1 value = [1, 0] class = 0

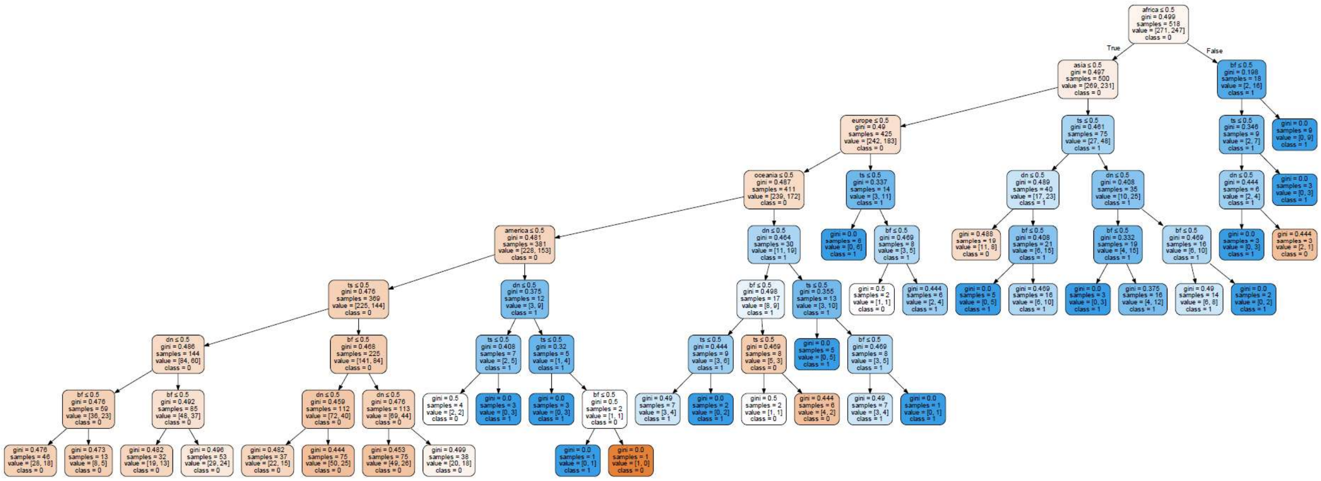
Zimbabwe


bf ≤ 0.5 gini = 0.499 samples = 424 value = [219, 205] class = 0
True
False
afr ≤ 0.5 gini = 0.498 samples = 234 value = [109, 125] class = 1
ts ≤ 0.5 gini = 0.488 samples = 190 value = [110, 80] class = 0
ts ≤ 0.5 gini = 0.5 samples = 221 value = [108, 113] class = 1
ts ≤ 0.5 gini = 0.142 samples = 13 value = [1, 12] class = 1
ocn ≤ 0.5 gini = 0.498 samples = 99 value = [53, 46] class = 0
ocn ≤ 0.5 gini = 0.468 samples = 91 value = [57, 34] class = 0
dn ≤ 0.5 gini = 0.495 samples = 98 value = [54, 44] class = 0
dn ≤ 0.5 gini = 0.493 samples = 123 value = [54, 69] class = 1
gini = 0.0 samples = 6 value = [0, 6] class = 1
gini = 0.245 samples = 7 value = [1, 6] class = 1
asa ≤ 0.5 gini = 0.493 samples = 82 value = [46, 36] class = 0
dn ≤ 0.5 gini = 0.484 samples = 17 value = [7, 10] class = 1
amr ≤ 0.5 gini = 0.476 samples = 82 value = [50, 32] class = 0
dn ≤ 0.5 gini = 0.346 samples = 9 value = [7, 2] class = 0
ocn ≤ 0.5 gini = 0.477 samples = 56 value = [34, 22] class = 0
ocn ≤ 0.5 gini = 0.499 samples = 42 value = [20, 22] class = 1
ocn ≤ 0.5 gini = 0.457 samples = 34 value = [12, 22] class = 1
amr ≤ 0.5 gini = 0.498 samples = 89 value = [42, 47] class = 1
dn ≤ 0.5 gini = 0.479 samples = 63 value = [38, 25] class = 0
dn ≤ 0.5 gini = 0.488 samples = 19 value = [8, 11] class = 1
gini = 0.375 samples = 4 value = [1, 3] class = 1
gini = 0.497 samples = 13 value = [6, 7] class = 1
asa ≤ 0.5 gini = 0.478 samples = 81 value = [49, 32] class = 0
gini = 0.0 samples = 1 value = [1, 0] class = 0
gini = 0.0 samples = 2 value = [2, 0] class = 0
gini = 0.408 samples = 7 value = [5, 2] class = 0
asa ≤ 0.5 gini = 0.473 samples = 47 value = [29, 18] class = 0
gini = 0.494 samples = 9 value = [5, 4] class = 0
asa ≤ 0.5 gini = 0.5 samples = 36 value = [18, 18] class = 0
gini = 0.444 samples = 6 value = [2, 4] class = 1
asa ≤ 0.5 gini = 0.42 samples = 30 value = [9, 21] class = 1
gini = 0.375 samples = 4 value = [3, 1] class = 0
ocn ≤ 0.5 gini = 0.5 samples = 78 value = [38, 40] class = 1
gini = 0.463 samples = 11 value = [4, 7] class = 1
gini = 0.43 samples = 16 value = [11, 5] class = 0
amr ≤ 0.5 gini = 0.489 samples = 47 value = [27, 20] class = 0
gini = 0.444 samples = 3 value = [1, 2] class = 1
gini = 0.492 samples = 16 value = [7, 9] class = 1
dn ≤ 0.5 gini = 0.484 samples = 68 value = [40, 28] class = 0
dn ≤ 0.5 gini = 0.426 samples = 13 value = [9, 4] class = 0
gini = 0.476 samples = 41 value = [25, 16] class = 0
gini = 0.444 samples = 6 value = [4, 2] class = 0
amr ≤ 0.5 gini = 0.497 samples = 24 value = [13, 11] class = 0
gini = 0.486 samples = 12 value = [5, 7] class = 1
amr ≤ 0.5 gini = 0.453 samples = 26 value = [9, 17] class = 1
gini = 0.0 samples = 4 value = [0, 4] class = 1
asa ≤ 0.5 gini = 0.499 samples = 58 value = [28, 30] class = 1
gini = 0.5 samples = 20 value = [10, 10] class = 0
afr ≤ 0.5 gini = 0.484 samples = 39 value = [23, 16] class = 0
gini = 0.5 samples = 8 value = [4, 4] class = 0
afr ≤ 0.5 gini = 0.476 samples = 41 value = [25, 16] class = 0
afr ≤ 0.5 gini = 0.494 samples = 27 value = [15, 12] class = 0
gini = 0.0 samples = 1 value = [1, 0] class = 0
gini = 0.444 samples = 12 value = [8, 4] class = 0
gini = 0.494 samples = 18 value = [10, 8] class = 0
gini = 0.5 samples = 6 value = [3, 3] class = 0
gini = 0.461 samples = 25 value = [9, 16] class = 1
gini = 0.0 samples = 1 value = [0, 1] class = 1
gini = 0.499 samples = 44 value = [21, 23] class = 1
gini = 0.5 samples = 14 value = [7, 7] class = 0
gini = 0.486 samples = 36 value = [21, 15] class = 0
gini = 0.444 samples = 3 value = [2, 1] class = 0
gini = 0.473 samples = 39 value = [24, 15] class = 0
gini = 0.5 samples = 2 value = [1, 1] class = 0
gini = 0.497 samples = 26 value = [14, 12] class = 0
gini = 0.0 samples = 1 value = [1, 0] class = 0

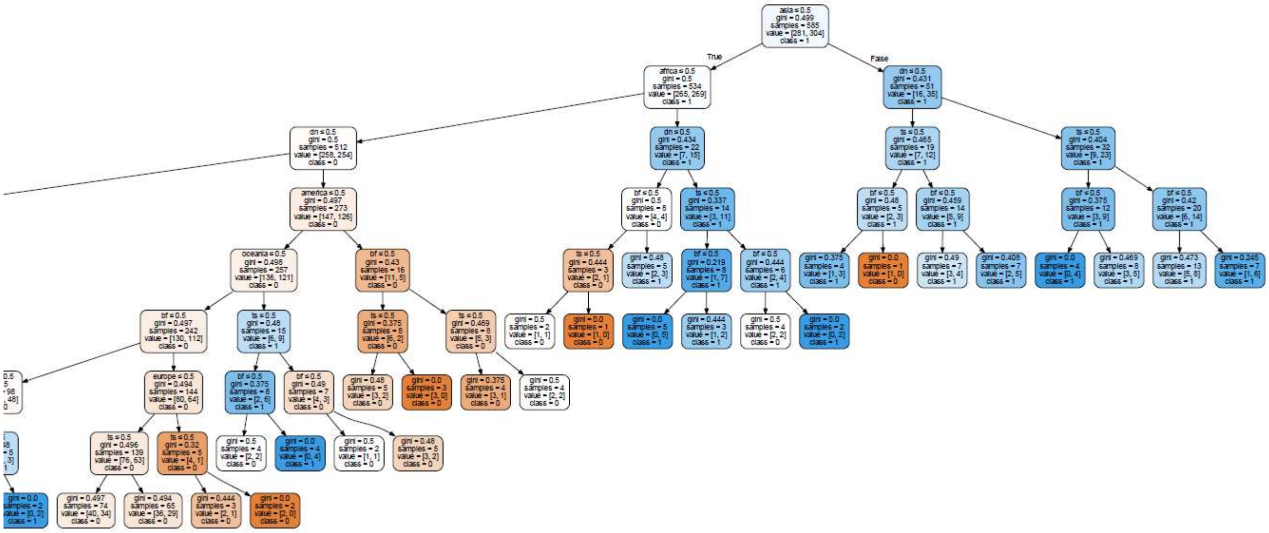
India


africa ≤ 0.5 gini = 0.5 samples = 608 value = [297, 311] class = 1
True
False
europe ≤ 0.5 gini = 0.5 samples = 569 value = [287, 282] class = 0
bf ≤ 0.5 gini = 0.381 samples = 39 value = [10, 29] class = 1
asia ≤ 0.5 gini = 0.499 samples = 541 value = [280, 261] class = 0
ts ≤ 0.5 gini = 0.375 samples = 28 value = [7, 21] class = 1
dn ≤ 0.5 gini = 0.266 samples = 19 value = [3, 16] class = 1
ts ≤ 0.5 gini = 0.455 samples = 20 value = [7, 13] class = 1
bf ≤ 0.5 gini = 0.485 samples = 58 value = [24, 34] class = 1
dn ≤ 0.5 gini = 0.49 samples = 14 value = [6, 8] class = 1
bf ≤ 0.5 gini = 0.133 samples = 14 value = [1, 13] class = 1
ts ≤ 0.5 gini = 0.219 samples = 8 value = [1, 7] class = 1
ts ≤ 0.5 gini = 0.298 samples = 11 value = [2, 9] class = 1
gini = 0.346 samples = 9 value = [2, 7] class = 1
dn ≤ 0.5 gini = 0.496 samples = 11 value = [5, 6] class = 1
ts ≤ 0.5 gini = 0.451 samples = 35 value = [12, 23] class = 1
dn ≤ 0.5 gini = 0.499 samples = 23 value = [12, 11] class = 0
bf ≤ 0.5 gini = 0.48 samples = 5 value = [2, 3] class = 1
bf ≤ 0.5 gini = 0.494 samples = 9 value = [4, 5] class = 1
gini = 0.0 samples = 10 value = [0, 10] class = 1
dn ≤ 0.5 gini = 0.375 samples = 4 value = [1, 3] class = 1
gini = 0.444 samples = 3 value = [1, 2] class = 1
gini = 0.0 samples = 5 value = [0, 5] class = 1
gini = 0.0 samples = 4 value = [0, 4] class = 1
gini = 0.408 samples = 7 value = [2, 5] class = 1
gini = 0.32 samples = 5 value = [1, 4] class = 1
gini = 0.444 samples = 6 value = [4, 2] class = 0
dn ≤ 0.5 gini = 0.43 samples = 16 value = [5, 11] class = 1
ts ≤ 0.5 gini = 0.469 samples = 8 value = [5, 3] class = 0
ts ≤ 0.5 gini = 0.498 samples = 15 value = [7, 8] class = 1
gini = 0.5 samples = 4 value = [2, 2] class = 0
gini = 0.0 samples = 1 value = [0, 1] class = 1
gini = 0.444 samples = 3 value = [1, 2] class = 1
gini = 0.5 samples = 6 value = [3, 3] class = 0
gini = 0.0 samples = 2 value = [0, 2] class = 1
gini = 0.5 samples = 2 value = [1, 1] class = 0
gini = 0.444 samples = 3 value = [2, 1] class = 0
gini = 0.355 samples = 13 value = [3, 10] class = 1
gini = 0.5 samples = 2 value = [1, 1] class = 0
gini = 0.444 samples = 6 value = [4, 2] class = 0
gini = 0.494 samples = 9 value = [4, 5] class = 1
gini = 0.5 samples = 6 value = [3, 3] class = 0

Sri Lanka


africa ≤ 0.5 gini = 0.499 samples = 518 value = [271, 247] class = 0
True
False
asia ≤ 0.5 gini = 0.497 samples = 500 value = [269, 231] class = 0
bf ≤ 0.5 gini = 0.198 samples = 18 value = [2, 16] class = 1
europe ≤ 0.5 gini = 0.49 samples = 425 value = [242, 183] class = 0
ts ≤ 0.5 gini = 0.461 samples = 75 value = [27, 48] class = 1
ts ≤ 0.5 gini = 0.346 samples = 9 value = [2, 7] class = 1
gini = 0.0 samples = 9 value = [0, 9] class = 1
oceania ≤ 0.5 gini = 0.487 samples = 411 value = [239, 172] class = 0
ts ≤ 0.5 gini = 0.337 samples = 14 value = [3, 11] class = 1
dn ≤ 0.5 gini = 0.489 samples = 40 value = [17, 23] class = 1
dn ≤ 0.5 gini = 0.408 samples = 35 value = [10, 25] class = 1
dn ≤ 0.5 gini = 0.444 samples = 6 value = [2, 4] class = 1
gini = 0.0 samples = 3 value = [0, 3] class = 1
america ≤ 0.5 gini = 0.481 samples = 381 value = [228, 153] class = 0
dn ≤ 0.5 gini = 0.464 samples = 30 value = [11, 19] class = 1
gini = 0.0 samples = 6 value = [0, 6] class = 1
bf ≤ 0.5 gini = 0.469 samples = 8 value = [3, 5] class = 1
gini = 0.488 samples = 19 value = [11, 8] class = 0
bf ≤ 0.5 gini = 0.408 samples = 21 value = [6, 15] class = 1
bf ≤ 0.5 gini = 0.332 samples = 19 value = [4, 15] class = 1
bf ≤ 0.5 gini = 0.469 samples = 16 value = [6, 10] class = 1
gini = 0.0 samples = 3 value = [0, 3] class = 1
gini = 0.444 samples = 3 value = [2, 1] class = 0
ts ≤ 0.5 gini = 0.476 samples = 369 value = [225, 144] class = 0
dn ≤ 0.5 gini = 0.375 samples = 12 value = [3, 9] class = 1
bf ≤ 0.5 gini = 0.498 samples = 17 value = [8, 9] class = 1
ts ≤ 0.5 gini = 0.355 samples = 13 value = [3, 10] class = 1
gini = 0.5 samples = 2 value = [1, 1] class = 0
gini = 0.444 samples = 6 value = [2, 4] class = 1
gini = 0.0 samples = 5 value = [0, 5] class = 1
gini = 0.469 samples = 16 value = [6, 10] class = 1
gini = 0.0 samples = 3 value = [0, 3] class = 1
gini = 0.375 samples = 16 value = [4, 12] class = 1
gini = 0.49 samples = 14 value = [6, 8] class = 1
gini = 0.0 samples = 2 value = [0, 2] class = 1
dn ≤ 0.5 gini = 0.486 samples = 144 value = [84, 60] class = 0
bf ≤ 0.5 gini = 0.468 samples = 225 value = [141, 84] class = 0
ts ≤ 0.5 gini = 0.408 samples = 7 value = [2, 5] class = 1
ts ≤ 0.5 gini = 0.32 samples = 5 value = [1, 4] class = 1
ts ≤ 0.5 gini = 0.444 samples = 9 value = [3, 6] class = 1
ts ≤ 0.5 gini = 0.469 samples = 8 value = [5, 3] class = 0
gini = 0.0 samples = 5 value = [0, 5] class = 1
bf ≤ 0.5 gini = 0.469 samples = 8 value = [3, 5] class = 1
bf ≤ 0.5 gini = 0.476 samples = 59 value = [36, 23] class = 0
bf ≤ 0.5 gini = 0.492 samples = 85 value = [48, 37] class = 0
dn ≤ 0.5 gini = 0.459 samples = 112 value = [72, 40] class = 0
dn ≤ 0.5 gini = 0.476 samples = 113 value = [69, 44] class = 0
gini = 0.5 samples = 4 value = [2, 2] class = 0
gini = 0.0 samples = 3 value = [0, 3] class = 1
gini = 0.0 samples = 3 value = [0, 3] class = 1
bf ≤ 0.5 gini = 0.5 samples = 2 value = [1, 1] class = 0
gini = 0.49 samples = 7 value = [3, 4] class = 1
gini = 0.0 samples = 2 value = [0, 2] class = 1
gini = 0.5 samples = 2 value = [1, 1] class = 0
gini = 0.444 samples = 6 value = [4, 2] class = 0
gini = 0.49 samples = 7 value = [3, 4] class = 1
gini = 0.0 samples = 1 value = [0, 1] class = 1
gini = 0.476 samples = 46 value = [28, 18] class = 0
gini = 0.473 samples = 13 value = [8, 5] class = 0
gini = 0.482 samples = 32 value = [19, 13] class = 0
gini = 0.496 samples = 53 value = [29, 24] class = 0
gini = 0.482 samples = 37 value = [22, 15] class = 0
gini = 0.444 samples = 75 value = [50, 25] class = 0
gini = 0.453 samples = 75 value = [49, 26] class = 0
gini = 0.499 samples = 38 value = [20, 18] class = 0
gini = 0.0 samples = 1 value = [0, 1] class = 1
gini = 0.0 samples = 1 value = [1, 0] class = 0

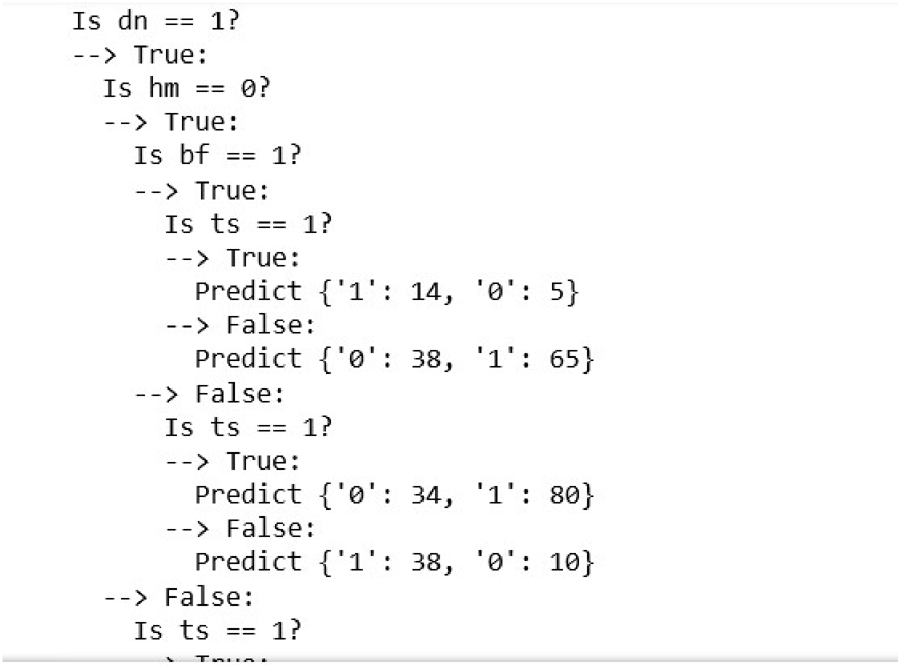
Pakistan


asia ≤ 0.5 gini = 0.499 samples = 585 value = [281, 304] class = 1
True
False
africa ≤ 0.5 gini = 0.5 samples = 534 value = [265, 269] class = 1
dn ≤ 0.5 gini = 0.431 samples = 51 value = [16, 35] class = 1
dn ≤ 0.5 gini = 0.5 samples = 512 value = [258, 254] class = 0
dn ≤ 0.5 gini = 0.434 samples = 22 value = [7, 15] class = 1
ts ≤ 0.5 gini = 0.465 samples = 19 value = [7, 12] class = 1
ts ≤ 0.5 gini = 0.404 samples = 32 value = [9, 23] class = 1
america ≤ 0.5 gini = 0.497 samples = 273 value = [147, 126] class = 0
bf ≤ 0.5 gini = 0.5 samples = 8 value = [4, 4] class = 0
ts ≤ 0.5 gini = 0.337 samples = 14 value = [3, 11] class = 1
bf ≤ 0.5 gini = 0.48 samples = 5 value = [2, 3] class = 1
bf ≤ 0.5 gini = 0.459 samples = 14 value = [5, 9] class = 1
bf ≤ 0.5 gini = 0.375 samples = 12 value = [3, 9] class = 1
bf ≤ 0.5 gini = 0.42 samples = 20 value = [6, 14] class = 1
oceania ≤ 0.5 gini = 0.498 samples = 257 value = [136, 121] class = 0
bf ≤ 0.5 gini = 0.43 samples = 16 value = [11, 5] class = 0
ts ≤ 0.5 gini = 0.444 samples = 3 value = [2, 1] class = 0
gini = 0.48 samples = 5 value = [2, 3] class = 1
bf ≤ 0.5 gini = 0.219 samples = 8 value = [1, 7] class = 1
bf ≤ 0.5 gini = 0.444 samples = 6 value = [2, 4] class = 1
gini = 0.375 samples = 4 value = [1, 3] class = 1
gini = 0.0 samples = 1 value = [1, 0] class = 0
gini = 0.49 samples = 7 value = [3, 4] class = 1
gini = 0.408 samples = 7 value = [2, 5] class = 1
gini = 0.0 samples = 4 value = [0, 4] class = 1
gini = 0.469 samples = 8 value = [3, 5] class = 1
gini = 0.473 samples = 13 value = [5, 8] class = 1
gini = 0.245 samples = 7 value = [1, 6] class = 1
bf ≤ 0.5 gini = 0.497 samples = 242 value = [130, 112] class = 0
ts ≤ 0.5 gini = 0.48 samples = 15 value = [6, 9] class = 1
ts ≤ 0.5 gini = 0.375 samples = 8 value = [6, 2] class = 0
ts ≤ 0.5 gini = 0.469 samples = 8 value = [5, 3] class = 0
gini = 0.5 samples = 2 value = [1, 1] class = 0
gini = 0.0 samples = 1 value = [1, 0] class = 0
gini = 0.0 samples = 5 value = [0, 5] class = 1
gini = 0.444 samples = 3 value = [1, 2] class = 1
gini = 0.5 samples = 4 value = [2, 2] class = 0
gini = 0.0 samples = 2 value = [0, 2] class = 1
0.5 5 98 48] 0
europe ≤ 0.5 gini = 0.494 samples = 144 value = [80, 64] class = 0
bf ≤ 0.5 gini = 0.375 samples = 8 value = [2, 6] class = 1
bf ≤ 0.5 gini = 0.49 samples = 7 value = [4, 3] class = 0
gini = 0.48 samples = 5 value = [3, 2] class = 0
gini = 0.0 samples = 3 value = [3, 0] class = 0
gini = 0.375 samples = 4 value = [3, 1] class = 0
gini = 0.5 samples = 4 value = [2, 2] class = 0
48 5 , 3] 1
ts ≤ 0.5 gini = 0.496 samples = 139 value = [76, 63] class = 0
ts ≤ 0.5 gini = 0.32 samples = 5 value = [4, 1] class = 0
gini = 0.5 samples = 4 value = [2, 2] class = 0
gini = 0.0 samples = 4 value = [0, 4] class = 1
gini = 0.5 samples = 2 value = [1, 1] class = 0
gini = 0.48 samples = 5 value = [3, 2] class = 0
gini = 0.0 samples = 2 value = [0, 2] class = 1
gini = 0.497 samples = 74 value = [40, 34] class = 0
gini = 0.494 samples = 65 value = [36, 29] class = 0
gini = 0.444 samples = 3 value = [2, 1] class = 0
gini = 0.0 samples = 2 value = [2, 0] class = 0

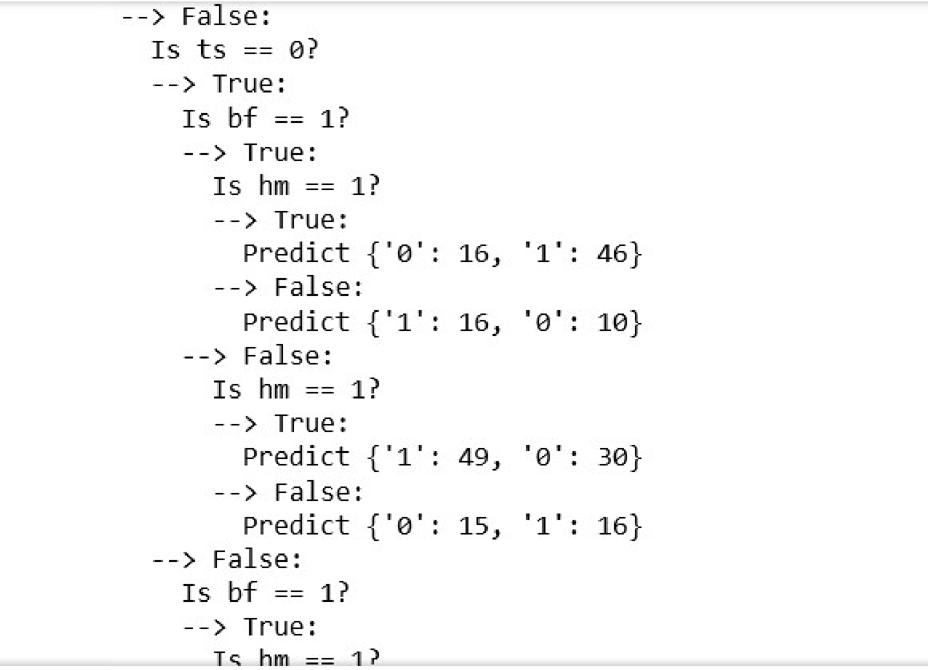
In the second approach we first design the questioning mechanism followed by the branching tree. It involves checking for valid numerical values followed by queries.

This is followed by partitioning the dataset into false rows and true rows depending on the query outcome. This leads to the formation of child nodes from the parent nodes. The gini impurity values for these rows are then calculated which further determines how the split will proceed.

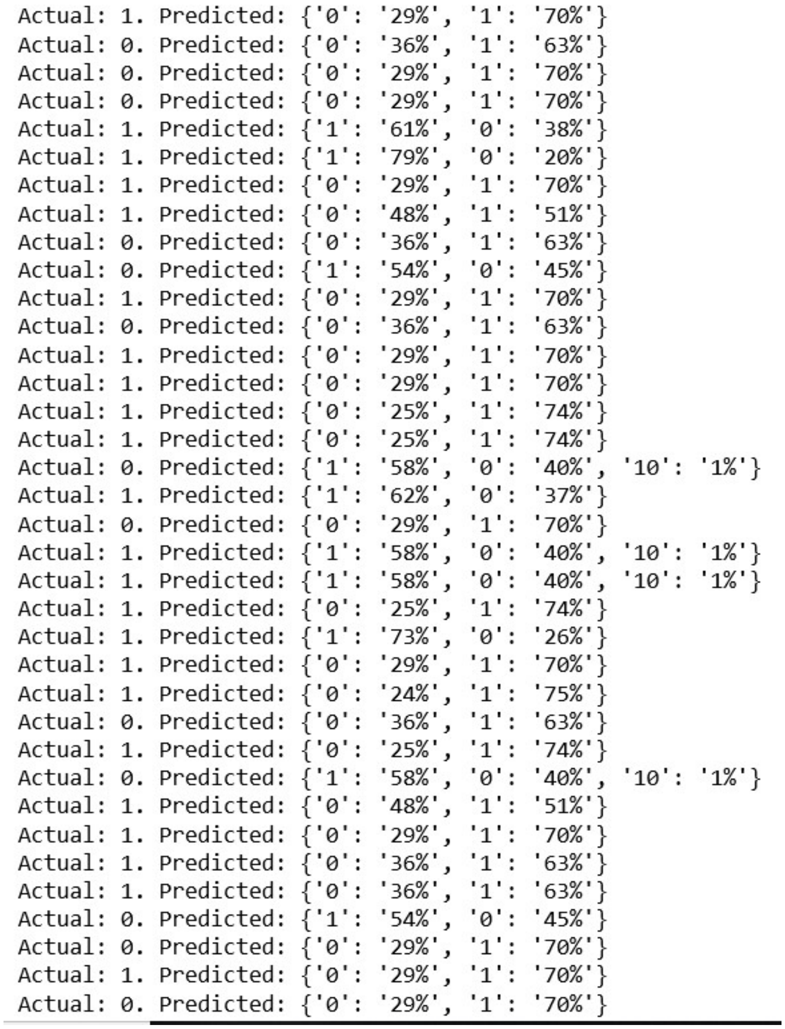
Amongst all the queries, the hierarchical order of optimal queries is decided and the tree structure is obtained.



```
Is dn == 1?
--> True:
  Is hm == 0?
  --> True:
    Is bf == 1?
    --> True:
      Is ts == 1?
      --> True:
        Predict {'1': 14, '0': 5}
      --> False:
        Predict {'0': 38, '1': 65}
    --> False:
      Is ts == 1?
      --> True:
        Predict {'0': 34, '1': 80}
      --> False:
        Predict {'1': 38, '0': 10}
  --> False:
    Is ts == 1?
    --> True:
```



```
    --> False:
      Is ts == 0?
      --> True:
        Is bf == 1?
        --> True:
          Is hm == 1?
          --> True:
            Predict {'0': 16, '1': 46}
          --> False:
            Predict {'1': 16, '0': 10}
        --> False:
          Is hm == 1?
          --> True:
            Predict {'1': 49, '0': 30}
          --> False:
            Predict {'0': 15, '1': 16}
      --> False:
        Is bf == 1?
        --> True:
          Is hm == 1?
```



```
Actual: 1. Predicted: {'0': '29%', '1': '70%'}
Actual: 0. Predicted: {'0': '36%', '1': '63%'}
Actual: 0. Predicted: {'0': '29%', '1': '70%'}
Actual: 0. Predicted: {'0': '29%', '1': '70%'}
Actual: 1. Predicted: {'1': '61%', '0': '38%'}
Actual: 1. Predicted: {'1': '79%', '0': '20%'}
Actual: 1. Predicted: {'0': '29%', '1': '70%'}
Actual: 1. Predicted: {'0': '48%', '1': '51%'}
Actual: 0. Predicted: {'0': '36%', '1': '63%'}
Actual: 0. Predicted: {'1': '54%', '0': '45%'}
Actual: 1. Predicted: {'0': '29%', '1': '70%'}
Actual: 0. Predicted: {'0': '36%', '1': '63%'}
Actual: 1. Predicted: {'0': '29%', '1': '70%'}
Actual: 1. Predicted: {'0': '29%', '1': '70%'}
Actual: 1. Predicted: {'0': '25%', '1': '74%'}
Actual: 1. Predicted: {'0': '25%', '1': '74%'}
Actual: 0. Predicted: {'1': '58%', '0': '40%', '10': '1%'}
Actual: 1. Predicted: {'1': '62%', '0': '37%'}
Actual: 0. Predicted: {'0': '29%', '1': '70%'}
Actual: 1. Predicted: {'1': '58%', '0': '40%', '10': '1%'}
Actual: 1. Predicted: {'1': '58%', '0': '40%', '10': '1%'}
Actual: 1. Predicted: {'0': '25%', '1': '74%'}
Actual: 1. Predicted: {'1': '73%', '0': '26%'}
Actual: 1. Predicted: {'0': '29%', '1': '70%'}
Actual: 1. Predicted: {'0': '24%', '1': '75%'}
Actual: 0. Predicted: {'0': '36%', '1': '63%'}
Actual: 1. Predicted: {'0': '25%', '1': '74%'}
Actual: 0. Predicted: {'1': '58%', '0': '40%', '10': '1%'}
Actual: 1. Predicted: {'0': '48%', '1': '51%'}
Actual: 1. Predicted: {'0': '29%', '1': '70%'}
Actual: 1. Predicted: {'0': '36%', '1': '63%'}
Actual: 1. Predicted: {'0': '36%', '1': '63%'}
Actual: 0. Predicted: {'1': '54%', '0': '45%'}
Actual: 0. Predicted: {'0': '29%', '1': '70%'}
Actual: 1. Predicted: {'0': '29%', '1': '70%'}
Actual: 0. Predicted: {'0': '29%', '1': '70%'}
```

CONCLUSION

TEAM	ACCURACY SCORE	
	Logistic regression	CART
India	57.1269	55.6976
Pakistan	54.2874	54.3551
Sri Lanka	52.8412	51.2545
Australia	65.8421	62.2587
South Africa	63.2587	58.7946
New Zealand	66.9846	64.1634
West Indies	50.8842	50.2146
England	64.7663	61.9845
Zimbabwe	72.0671	69.7945

We conclude that logistic regression achieves a higher accuracy in contrast to CART approach but CART is easier for interpretation and explanation.

****