

INDIANA UNIVERISTY BLOOMINGTON



PA 1: Decision Trees

I 526 Applied Machine Learning

SHORT REPORT

Submitted by

Nihar Khetan

Masters Candidate in Computer Science
School of Informatics and Computing

Indiana University

nkhetan@indiana.edu

Ghanshyam Malu

Masters Candidate in Computer Science
School of Informatics and Computing

Indiana University

gmalu@indiana.edu

Under the Guidance of:

Professor Sriraam Natarajan

Asst. Professor, School of Informatics and
Computing

Indiana University

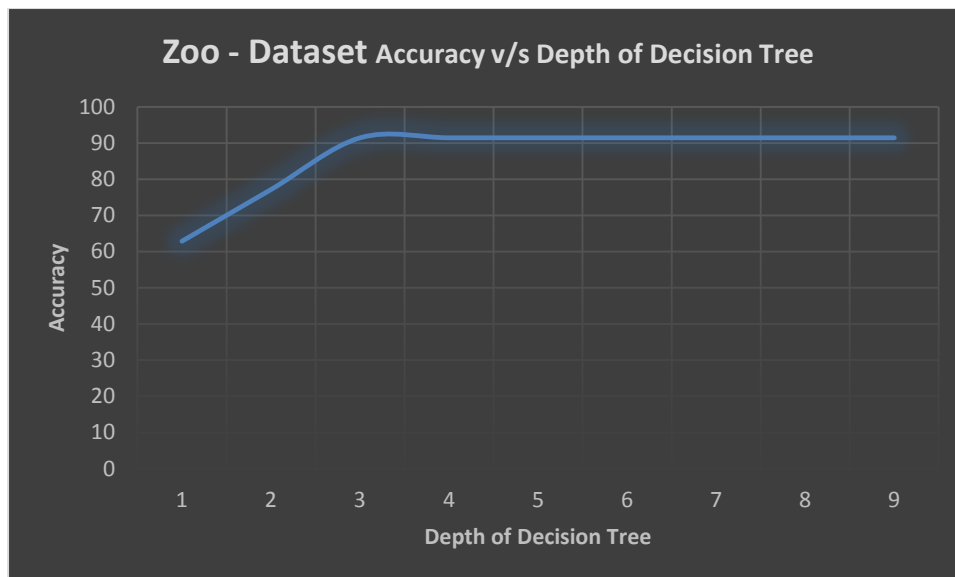
Contents

1. Depth Analysis – Zoo dataset.....	2
2. Decision Tree and Confusion Matrix – Zoo Dataset.....	3
1.1 Depth = 1	3
1.1 Depth = 2	4
1.1 Depth = 3	5
3. Evaluation with Weka – Zoo Dataset.....	7
4. Decision Tree and Confusion Matrix – Carvana Dataset.....	8
4.1 Depth = 1	8
4.2 Depth = 2	9
5. Evaluation with Weka – Carvana Dataset.....	11

1. Depth Analysis – Zoo dataset

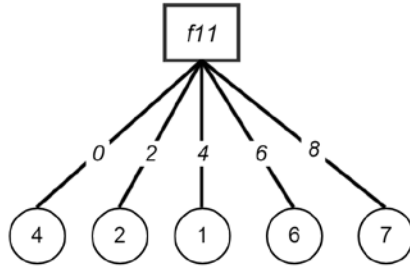
The Decision Tree modeler was run against the provided zoo dataset at various depths shown below. It can be seen that the performance of model is considerably bad at depth 1 and depth 2. At such low depth restriction, the decision tree usually cannot provide highly accurate results. This is due to the classification based on the highest appearing class label in the training dataset.

Zoo – Dataset (Decision Tree)		
Depth	Accuracy	Error Rate
1	62.86	37.14
2	77.14	22.86
3	91.43	8.57
4	91.43	8.57
5	91.43	8.57
6	91.43	8.57
7	91.43	8.57
8	91.43	8.57
9	91.43	8.57



2. Decision Tree and Confusion Matrix – Zoo Dataset

1.1 Depth = 1



```

*****
                        Decision Tree
*****

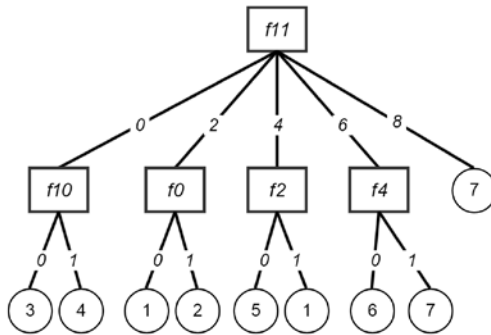
-----
Attribute: root      Node: f11      Level: 0      Children Count: 5
-----

***Children***
Attribute: 0  Label : 4  Level: 1      ClassificationLabel: 4 ***** Label *****
Attribute: 2  Label : 2  Level: 1      ClassificationLabel: 2 ***** Label *****
Attribute: 4  Label : 1  Level: 1      ClassificationLabel: 1 ***** Label *****
Attribute: 6  Label : 6  Level: 1      ClassificationLabel: 6 ***** Label *****
Attribute: 8  Label : 7  Level: 1      ClassificationLabel: 7 ***** Label *****
  
```

Confusion Matrix Depth=1

	Predicted Value							
		1	3	2	5	4	7	6
Expected Values	1	8	0	4	0	2	0	0
	3	2	0	0	0	0	0	0
	2	0	0	7	0	0	0	0
	5	1	0	0	0	0	0	0
	4	0	0	0	0	4	0	0
	7	1	0	0	0	3	0	0
	6	0	0	0	0	0	0	3

1.1 Depth = 2



```

*****
Decision Tree
*****

-----
Attribute: root Feature: f11 Level: 0 Children Count: 5
-----

***Children***
Attribute: 0 Feature : f10 Level: 1
Attribute: 2 Feature : f0 Level: 1
Attribute: 4 Feature : f2 Level: 1
Attribute: 6 Feature : f4 Level: 1
Attribute: 8 Label : 7 Level: 1 ClassificationLabel: 7 ***** Label *****

-----
Attribute: 0 Feature: f10 Level: 1 Parent Feature: f11 Children Count: 2
-----

***Children***
Attribute: 0 Label : 3 Level: 2 ClassificationLabel: 3 ***** Label *****
Attribute: 1 Label : 4 Level: 2 ClassificationLabel: 4 ***** Label *****

-----
Attribute: 2 Feature: f0 Level: 1 Parent Feature: f11 Children Count: 2
-----

***Children***
Attribute: 0 Label : 1 Level: 2 ClassificationLabel: 1 ***** Label *****
Attribute: 1 Label : 2 Level: 2 ClassificationLabel: 2 ***** Label *****

-----
Attribute: 4 Feature: f2 Level: 1 Parent Feature: f11 Children Count: 2
-----

***Children***
Attribute: 0 Label : Level: 2 ClassificationLabel: 5 ***** Label *****
Attribute: 1 Label : Level: 2 ClassificationLabel: 1 ***** Label *****

-----
Attribute: 6 Feature: f4 Level: 1 Parent Feature: f11 Children Count: 2
-----

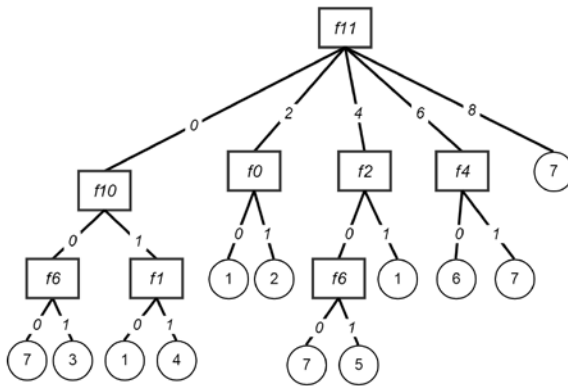
***Children***
Attribute: 0 Label : 6 Level: 2 ClassificationLabel: 6 ***** Label *****
Attribute: 1 Label : 7 Level: 2 ClassificationLabel: 7 ***** Label *****

```

Confusion Matrix Depth=2

		Predicted Value						
Expected Values		1	3	2	5	4	7	6
	1	12	0	0	0	2	0	0
	3	0	0	0	2	0	0	0
	2	0	0	7	0	0	0	0
	5	0	0	0	1	0	0	0
	4	0	0	0	0	4	0	0
	7	1	3	0	0	0	0	0
	6	0	0	0	0	0	0	3

1.1 Depth = 3



 Decision Tree

 Attribute: root Node: f11 Level: 0 Children Count: 5

Children

Attribute: 0 Node: f10 Level: 1
 Attribute: 2 Node: f0 Level: 1
 Attribute: 4 Node: f2 Level: 1
 Attribute: 6 Node: f4 Level: 1
 Attribute: 8 Node: 7 Level: 1 ClassificationLabel: 7 ***** Leaf *****

 Attribute: 0 Node: f10 Level: 1 Children Count: 2

Children

Attribute: 0 Node: f6 Level: 2
 Attribute: 1 Node: f1 Level: 2

```

Attribute: 0      Node: f6      Level: 2      Children Count: 2
-----
***Children***
Attribute: 0      Node: 7 Level: 3      ClassificationLabel: 7 ***** Leaf *****
Attribute: 1      Node: 3 Level: 3      ClassificationLabel: 3 ***** Leaf *****
-----
Attribute: 1      Node: f1      Level: 2      Children Count: 2
-----
***Children***
Attribute: 0      Node: 1 Level: 3      ClassificationLabel: 1 ***** Leaf *****
Attribute: 1      Node: 4 Level: 3      ClassificationLabel: 4 ***** Leaf *****
-----
Attribute: 2      Node: f0      Level: 1      Children Count: 2
-----
***Children***
Attribute: 0      Node: 1 Level: 2      ClassificationLabel: 1 ***** Leaf *****
Attribute: 1      Node: 2 Level: 2      ClassificationLabel: 2 ***** Leaf *****
-----
Attribute: 4      Node: f2      Level: 1      Children Count: 2
-----
***Children***
Attribute: 0      Node: f6      Level: 2
Attribute: 1      Node: 1 Level: 2      ClassificationLabel: 1 ***** Leaf *****
-----
Attribute: 0      Node: f6      Level: 2      Children Count: 2
-----
***Children***
Attribute: 0      Node: 7 Level: 3      ClassificationLabel: 7 ***** Leaf *****
Attribute: 1      Node: 5 Level: 3      ClassificationLabel: 5 ***** Leaf *****
-----
Attribute: 6      Node: f4      Level: 1      Children Count: 2
-----
***Children***
Attribute: 0      Node: 6 Level: 2      ClassificationLabel: 6 ***** Leaf *****
Attribute: 1      Node: 7 Level: 2      ClassificationLabel: 7 ***** Leaf *****

```

Confusion Matrix Depth=3

	Predicted Value							
Expected Values		1	3	2	5	4	7	6
	1	14	0	0	0	0	0	0
	3	0	0	0	1	0	1	0
	2	0	0	7	0	0	0	0
	5	0	0	0	1	0	0	0
	4	0	0	0	0	4	0	0
	7	1	0	0	0	0	3	0
	6	0	0	0	0	0	0	3

3. Evaluation with Weka – Zoo Dataset

Following default J48 algorithm in weka

=== Summary ===

Correctly Classified Instances	32	91.4286 %
Incorrectly Classified Instances	3	8.5714 %

Decision tree built

J48 pruned tree

```
f0 = 0
|   f2 = 0
|   |   f6 = 0
|   |   |   f3 = 0: 7 (7.0/1.0)
|   |   |   f3 = 1: 6 (4.0)
|   |   f6 = 1
|   |   |   f10 = 0
|   |   |   |   f11 = 0: 3 (3.0)
|   |   |   |   f11 = 2: 3 (0.0)
|   |   |   |   f11 = 4: 5 (3.0)
|   |   |   |   f11 = 6: 3 (0.0)
|   |   |   |   f11 = 8: 3 (0.0)
|   |   |   f10 = 1: 4 (9.0)
|   f2 = 1: 1 (27.0)
f0 = 1: 2 (13.0)
```

Number of Leaves : 10

Size of the tree : 16

Confusion Matrix

=== Confusion Matrix ===

```

  a  b  c  d  e  f  g  <-- classified as
14  0  0  0  0  0  0 | a = 1
 0  7  0  0  0  0  0 | b = 2
 0  0  0  0  1  0  1 | c = 3
 0  0  0  4  0  0  0 | d = 4
 0  0  0  0  1  0  0 | e = 5
 0  0  0  0  0  2  1 | f = 6
 0  0  0  0  0  0  4 | g = 7

```

4. Decision Tree and Confusion Matrix – Carvana Dataset

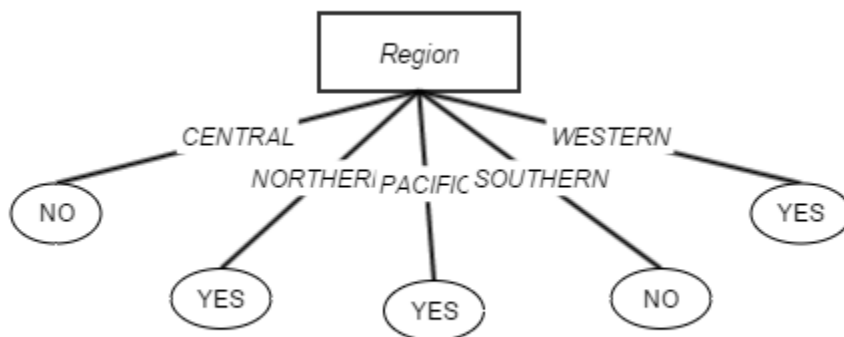
A used car business dealer does his business by purchasing used cars and selling them to potential buyers. He can suffer a substantial loss if the car he purchases has any serious defect which might prevent it from being sold. This can happen due to many reasons like defective machineries or other issues in the cars. We are trying to predict if a car purchased by a dealer is a good/bad purchase. Our aim is to predict whether a car purchased at an auction is a **Good Buy** or a **Bad Buy**.

The classifier label is `IsBadBuy` with values as either **Yes** or **No**.

Training set = 1000 data points (500 good buys, 500 bad buys)

Test set = 75 data points (25 bad buys, 50 good buys)

4.1 Depth = 1



```

*****
Decision Tree
*****

-----
Attribute: root      Node: Auction      Level: 0      Children Count: 3
-----

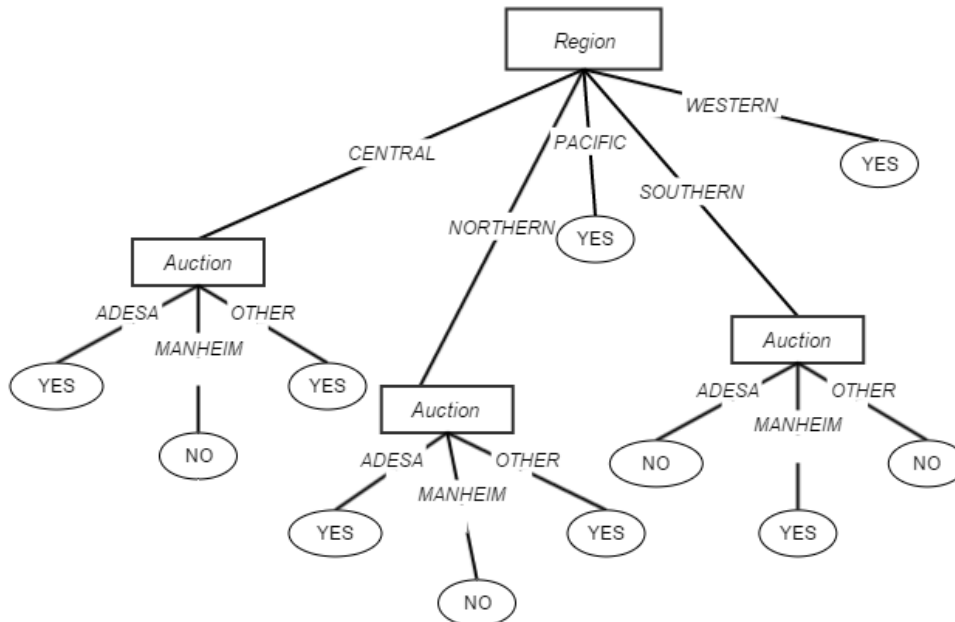
***Children***
Attribute: ADESA      Node: No      Level: 1      ClassificationLabel: No      ***** Leaf *****
Attribute: MANHEIM    Node: Yes     Level: 1      ClassificationLabel: Yes     ***** Leaf *****
Attribute: OTHER      Node: Yes     Level: 1      ClassificationLabel: Yes     ***** Leaf *****
    
```

Confusion Matrix Depth=1

Accuracy = 68%

Expected Values	Predicted Value		
		Yes	No
	Yes	0	24
	No	0	51

4.2 Depth = 2



Decision Tree

Attribute: root Feature: Region Level: 0 Children Count: 5

Children

Attribute: CENTRAL Feature : Auction Level: 1
Attribute: NORTHERN Feature : Auction Level: 1
Attribute: PACIFIC Label : Yes Level: 1 ClassificationLabel: Yes Label
Attribute: SOUTHERN Feature : Auction Level: 1
Attribute: WESTERN Label : Yes Level: 1 ClassificationLabel: Yes Label

Attribute: CENTRAL Feature: Auction Level: 1 Parent Feature: Region Children Count: 3

Children

Attribute: ADESA Label : Yes Level: 2 ClassificationLabel: Yes Label
Attribute: MANHEIM Label : No Level: 2 ClassificationLabel: No Label
Attribute: OTHER Label : Yes Level: 2 ClassificationLabel: Yes Label

Attribute: NORTHERN Feature: Auction Level: 1 Parent Feature: Region Children Count: 3

Children

Attribute: ADESA Label : Yes Level: 2 ClassificationLabel: Yes Label
Attribute: MANHEIM Label : No Level: 2 ClassificationLabel: No Label
Attribute: OTHER Label : Yes Level: 2 ClassificationLabel: Yes Label

Attribute: SOUTHERN Feature: Auction Level: 1 Parent Feature: Region Children Count: 3

Children

Attribute: ADESA Label : No Level: 2 ClassificationLabel: No Label
Attribute: MANHEIM Label : Yes Level: 2 ClassificationLabel: Yes Label
Attribute: OTHER Label : No Level: 2 ClassificationLabel: No Label

Confusion Matrix Depth=2

Accuracy: 73.33%

Expected Values	Predicted Value		
		Yes	No
	Yes	4	20
	No	0	51

Accuracy at different depths

Carvana – Dataset (Decision Tree)		
Depth	Accuracy	%
1	68	
2	73.33	
3	69.33	
4	73.33	
5	72	
6	81.33	
7	84	
8	82.7	
9	85.33	
10	86.7	
11	86.7	
12	86.7	

5. Evaluation with Weka – Carvana Dataset

Please scroll down to see the big tree.

Correctly Classified Instances	61	81.3333 %
Incorrectly Classified Instances	14	18.6667 %

J48 pruned tree

Region = SOUTHERN

```

|   Auction = ADESA
|   |   AvgConditionRetailPrice = >=7162 and <8498
|   |   |   AvgConditionPrice = >=3300 and <4447: No (20.0/5.0)
|   |   |   AvgConditionPrice = >=2440 and <3300: No (0.0)
|   |   |   AvgConditionPrice = <2440: No (0.0)
|   |   |   AvgConditionPrice = >=4447 and <6529: No (32.0/10.0)
|   |   |   AvgConditionPrice = >=6529 and <18176: Yes (3.0)
|   |   |   AvgConditionPrice = >18176: No (0.0)
|   |   AvgConditionRetailPrice = >=4067 and <7162
|   |   |   AvgConditionPrice = >=3300 and <4447
|   |   |   |   AvgConditionCurrentDayPrice = >=4776 and <6650: Yes (2.0/1.0)
|   |   |   |   AvgConditionCurrentDayPrice = >=2743 and <3551: Yes (16.0/5.0)
|   |   |   |   AvgConditionCurrentDayPrice = <2743: No (1.0)
|   |   |   |   AvgConditionCurrentDayPrice = >=6650 and <18502: Yes (0.0)
|   |   |   |   AvgConditionCurrentDayPrice = >=3551 and <4776
|   |   |   |   |   Size = LARGE
|   |   |   |   |   |   VehOdo = >=73550 and <89600: Yes (2.0)
|   |   |   |   |   |   VehOdo = <56980: Yes (0.0)
|   |   |   |   |   |   VehOdo = >=56980 and <73550: No (2.0)
|   |   |   |   |   |   VehOdo = >89600: Yes (0.0)
|   |   |   |   |   Size = MEDIUM
|   |   |   |   |   |   VehOdo = >=73550 and <89600: No (3.0)
|   |   |   |   |   |   VehOdo = <56980: No (2.0)
|   |   |   |   |   |   VehOdo = >=56980 and <73550
|   |   |   |   |   |   |   VehicleAge = FOUR: No (2.0)
|   |   |   |   |   |   |   VehicleAge = FIVE: Yes (2.0)
|   |   |   |   |   |   |   VehicleAge = MoreThanSIX: Yes (0.0)
|   |   |   |   |   |   |   VehicleAge = THREE: Yes (0.0)
|   |   |   |   |   |   |   VehicleAge = Upto 2: Yes (0.0)
|   |   |   |   |   |   |   VehOdo = >89600: Yes (2.0)
|   |   |   |   |   Size = SMALL: Yes (2.0)
|   |   |   |   |   Size = SPECIAL: Yes (1.0)
|   |   |   |   AvgConditionCurrentDayPrice = >18502: Yes (0.0)
|   |   |   AvgConditionPrice = >=2440 and <3300
|   |   |   VehOdo = >=73550 and <89600: No (16.0/4.0)

```

```

| | | | VehOdo = >=73550 and <89600: No (16.0/4.0)
| | | | VehOdo = <56980: Yes (5.0/1.0)
| | | | VehOdo = >=56980 and <73550: No (10.0/2.0)
| | | | VehOdo = >89600
| | | | | WheelType = Alloy: Yes (4.0)
| | | | | WheelType = Covers: No (3.0/1.0)
| | | | | WheelType = Special: Yes (0.0)
| | | | AvgConditionPrice = <2440: No (11.0/3.0)
| | | | AvgConditionPrice = >=4447 and <6529: Yes (13.0)
| | | | AvgConditionPrice = >=6529 and <18176: Yes (0.0)
| | | | AvgConditionPrice = >18176: Yes (0.0)
| | | AvgConditionRetailPrice = >=8498 and <22710
| | | | AvgConditionCurrentDayPrice = >=4776 and <6650
| | | | | VehOdo = >=73550 and <89600
| | | | | | Transmission = AUTO: Yes (19.0/5.0)
| | | | | | Transmission = MANUAL: No (3.0)
| | | | | VehOdo = <56980: No (15.0/3.0)
| | | | | VehOdo = >=56980 and <73550
| | | | | | Size = LARGE: Yes (7.0/1.0)
| | | | | | Size = MEDIUM: No (15.0/3.0)
| | | | | | Size = SMALL: Yes (2.0/1.0)
| | | | | | Size = SPECIAL: No (0.0)
| | | | | VehOdo = >89600
| | | | | | VehicleAge = FOUR: Yes (2.0)
| | | | | | VehicleAge = FIVE: No (3.0)
| | | | | | VehicleAge = MoreThanSIX: No (3.0/1.0)
| | | | | | VehicleAge = THREE: No (0.0)
| | | | | | VehicleAge = Upto 2: No (0.0)
| | | | AvgConditionCurrentDayPrice = >=2743 and <3551: No (0.0)
| | | | AvgConditionCurrentDayPrice = <2743: No (0.0)
| | | | AvgConditionCurrentDayPrice = >=6650 and <18502: No (97.0/23.0)
| | | | AvgConditionCurrentDayPrice = >=3551 and <4776: No (3.0)
| | | | AvgConditionCurrentDayPrice = >18502: No (0.0)
| | | AvgConditionRetailPrice = <4067: Yes (20.0)
| | | AvgConditionRetailPrice = >22710: No (0.0)
| | Auction = MANHEIM: Yes (29.0)
| | Auction = OTHER
| | | VehicleAge = FOUR
| | | | AvgConditionRetailPrice = >=7162 and <8498

```

```
| | | AvgConditionRetailPrice = >=7162 and <8498
| | | | AvgConditionCurrentDayPrice = >=4776 and <6650: No (2.0)
| | | | AvgConditionCurrentDayPrice = >=2743 and <3551: Yes (0.0)
| | | | AvgConditionCurrentDayPrice = <2743: Yes (0.0)
| | | | AvgConditionCurrentDayPrice = >=6650 and <18502: Yes (0.0)
| | | | AvgConditionCurrentDayPrice = >=3551 and <4776: Yes (5.0/1.0)
| | | | AvgConditionCurrentDayPrice = >18502: Yes (0.0)
| | | AvgConditionRetailPrice = >=4067 and <7162: No (12.0/1.0)
| | | AvgConditionRetailPrice = >=8498 and <22710: Yes (8.0/2.0)
| | | AvgConditionRetailPrice = <4067: No (0.0)
| | | AvgConditionRetailPrice = >22710: No (0.0)
| | VehicleAge = FIVE
| | | AvgConditionRetailPrice = >=7162 and <8498: Yes (3.0/1.0)
| | | AvgConditionRetailPrice = >=4067 and <7162: No (6.0/1.0)
| | | AvgConditionRetailPrice = >=8498 and <22710: Yes (7.0/2.0)
| | | AvgConditionRetailPrice = <4067: No (3.0/1.0)
| | | AvgConditionRetailPrice = >22710: No (0.0)
| | VehicleAge = MoreThanSIX
| | | AvgConditionRetailPrice = >=7162 and <8498: Yes (5.0)
| | | AvgConditionRetailPrice = >=4067 and <7162
| | | | WheelType = Alloy: No (4.0)
| | | | WheelType = Covers: Yes (3.0)
| | | | WheelType = Special: No (0.0)
| | | AvgConditionRetailPrice = >=8498 and <22710: Yes (4.0)
| | | AvgConditionRetailPrice = <4067: Yes (2.0/1.0)
| | | AvgConditionRetailPrice = >22710: Yes (0.0)
| | VehicleAge = THREE: No (49.0/4.0)
| | VehicleAge = Upto 2: No (25.0/1.0)
Region = CENTRAL
| Auction = ADESA: Yes (47.0)
| Auction = MANHEIM: No (159.0/14.0)
| Auction = OTHER: Yes (15.0)
Region = NORTHERN
| Auction = ADESA: Yes (10.0)
| Auction = MANHEIM: No (15.0/2.0)
| Auction = OTHER: Yes (2.0)
Region = PACIFIC: Yes (197.0/42.0)
Region = WESTERN: Yes (45.0)
```

Confusion Matrix

```
=== Confusion Matrix ===

  a  b  <-- classified as
16  8  |  a = Yes
 6 45  |  b = No
```