Driver Alertness (Implemented in R)

CSE675 – Stochastic Simulation and Inference

Karthik Kiran (kkiran@buffalo.edu)

Under the Guidance of:
Prof. Wen Dong

INTRODUCTION

- Driving while distracted or tired may lead to accidents
- Distractions can arise from – the driver, the vehicle or the environment
- Main goals of the project
  - To build a detector/ classifier to detect if the driver is alert or not
  - To solve the ever trending problem in the automobile safety sphere
  - To benefit insurance agencies and automotive safety industries by implementing a supervised learning algorithm.

DATASET

- Train Dataset

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
1	TrialID	ObsNum	IsAlert	P1	P2	P3	P4	P5	P6	P7	P8	E1	E2	E3	E4	E5	E6	E7	E8	E9	E10	E11	V1	V2	V3	V4
2	0	0	0	34.7406	9.84593	1400	42.8571	0.290601	572	104.895	0	0	0	0	1	-20	0.015875	324	1	1	1	57	0	101.96	0.175	752	5.99375
3	0	1	0	34.4215	13.4112	1400	42.8571	0.290601	572	104.895	0	0	0	0	1	-20	0.015875	324	1	1	1	57	0	101.98	0.455	752	5.99375
4	0	2	0	34.3447	15.1852	1400	42.8571	0.290601	576	104.167	0	0	0	0	1	-20	0.015875	324	1	1	1	57	0	101.97	0.28	752	5.99375
5	0	3	0	34.3421	8.84696	1400	42.8571	0.290601	576	104.167	0	0	0	0	1	-20	0.015875	324	1	1	1	57	0	101.99	0.07	752	5.99375
6	0	4	0	34.3322	14.6994	1400	42.8571	0.290601	576	104.167	0	0	0	0	1	-20	0.015875	324	1	1	1	57	0	102.07	0.175	752	5.99375
7	0	5	0	34.3729	13.6444	1400	42.8571	0.290601	576	104.167	0	0	0	0	1	-20	0.015875	324	1	1	1	57	0	102	0.28	752	5.99375
8	0	6	0	34.3851	10.1654	1400	42.8571	0.290601	576	104.167	0	0	0	0	1	-20	0.015875	324	1	1	1	57	0	102.01	0.175	752	5.99375
9	0	7	0	34.4313	13.327	1400	42.8571	0.290601	576	104.167	0	0	0	0	1	-20	0.015875	324	1	1	1	57	0	101.93	0.28	752	5.99375
10	0	8	0	34.497	17.7124	1400	42.8571	0.290601	592	101.351	0	0	0	0	1	-20	0.015875	324	1	1	1	57	0	101.88	0.07	752	5.99375
11	0	9	0	34.5994	6.88282	1400	42.8571	0.290601	592	101.351	0	0	0	0	1	-20	0.015875	324	1	1	1	57	0	101.93	0.175	752	5.99375
12	0	10	0	34.732	15.2195	1400	42.8571	0.290601	592	101.351	0	0	0	0	1	-20	0.015875	324	1	1	1	57	0	101.91	0	752	5.99375
13	0	11	0	35.0161	13.1415	1400	42.8571	0.290601	592	101.351	0	0	0	0	1	-20	0.015875	324	1	1	1	57	0	102	0	752	5.99375
14	0	12	0	35.3514	10.582	1400	42.8571	0.290601	592	101.351	0	0	0	0	1	-20	0.015875	324	1	1	1	57	0	101.86	0.175	752	5.99375
15	0	13	0	35.8153	13.4948	1400	42.8571	0.290601	592	101.351	0	0	0	0	1	-20	0.015875	324	1	1	1	57	0	101.76	0.455	752	7.48125

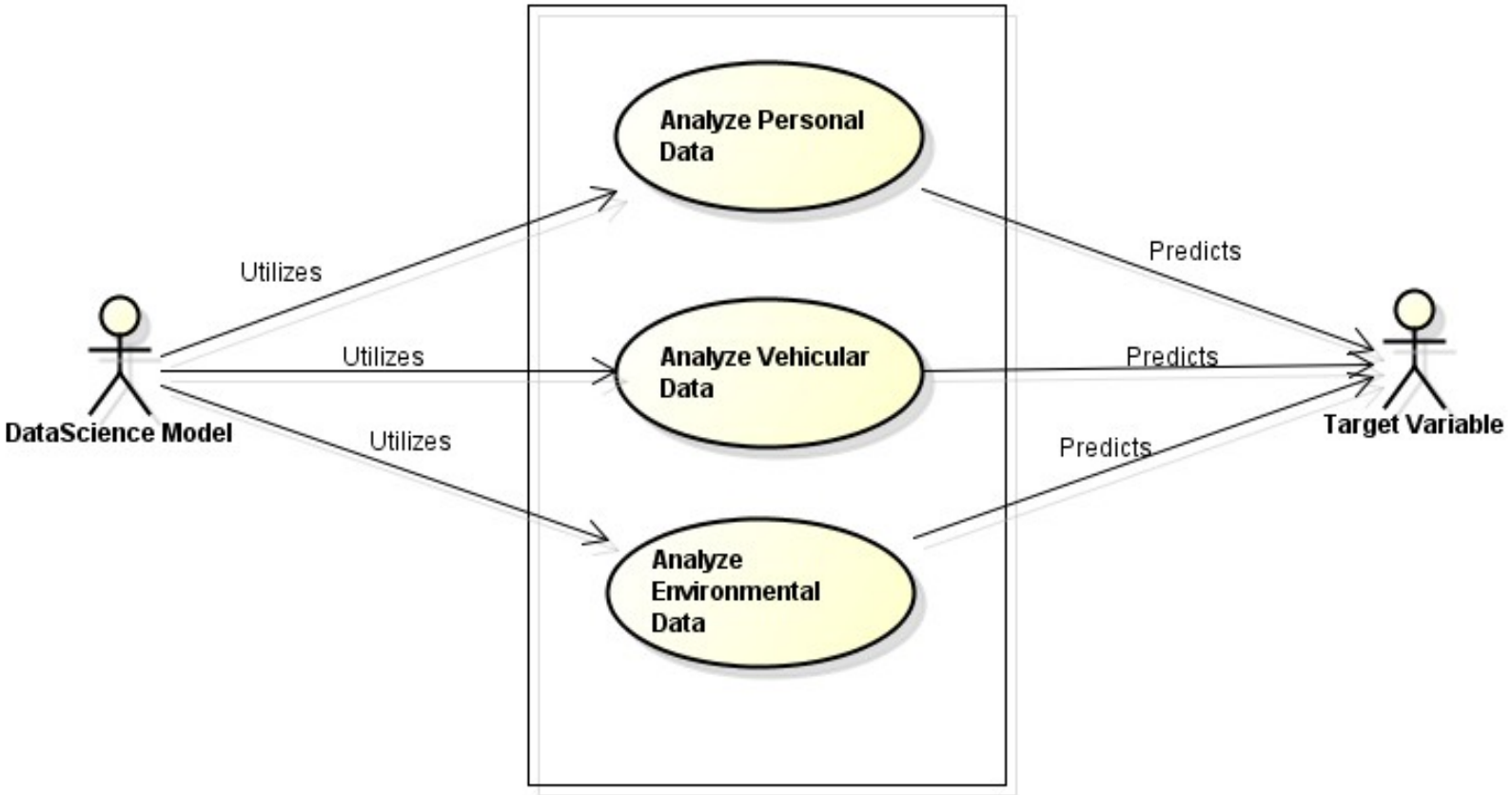
- Column 1 – Trial ID
- Column 2 – Observation Number
- Column 3 – Target Variable (IsAlert = 0 –> Driver is not alert, 1 –> Driver is Alert)
- Columns 4 to 12 – P1 through P8 (Physiological Data)
- Columns 13 to 24 – E1 through E11 (Environmental Data)
- Columns 25 to 36 – V1 through V11 (Vehicular Data)

DATASET (contd.)

- Test Dataset

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
1	TrialID	ObsNum	IsAlert	P1	P2	P3	P4	P5	P6	P7	P8	E1	E2	E3	E4	E5	E6	E7	E8	E9
2	0	0	?	38.4294	10.9435	1000	60	0.302277	508	118.11	0	0	0	4	4	0.015434	328	1	1	1
3	0	1	?	38.3809	15.3212	1000	60	0.302277	508	118.11	0	0	0	4	4	0.015434	328	1	1	1
4	0	2	?	38.2342	11.514	1000	60	0.302277	508	118.11	0	0	0	4	8	0.015938	328	1	1	1
5	0	3	?	37.9304	12.2615	1000	60	0.302277	508	118.11	0	0	0	4	8	0.015938	328	1	1	1
6	0	4	?	37.8085	12.3666	1000	60	0.302277	504	119.048	0	0	0	4	8	0.015938	328	1	1	1
7	0	5	?	37.7257	11.4631	1000	60	0.302277	504	119.048	0	0	0	4	8	0.015938	328	1	1	1
8	0	6	?	37.4322	14.8201	1000	60	0.302277	504	119.048	0	0	0	4	8	0.015938	328	1	1	1
9	0	7	?	37.3015	11.3936	1000	60	0.302277	504	119.048	0	0	0	4	8	0.015938	328	1	1	1

USE CASE DIAGRAM



PROPOSED SOLUTION

- Supervised learning algorithm to predict a binary target variable (0 or 1)

- Classifier can be run upon the dataset to predict target variable

- Techniques employed for classification:
  - Decision tree
  - Random forest

- Techniques tested independent of each other to evaluate precision and performance

FEATURE SELECTION

- Large number of features in the training dataset give inaccurate results.
- Selecting features which give maximum information – to ensure correlated features were not used during training of models.
- Resulted in faster training times and efficient output of prediction on testing dataset.
- Features like V11, E9 and E5 contribute a large extent to prediction of the target variable.

Decision Tree Model

- A decision tree graphically describes:

  Decisions to be made

  The events that may occur

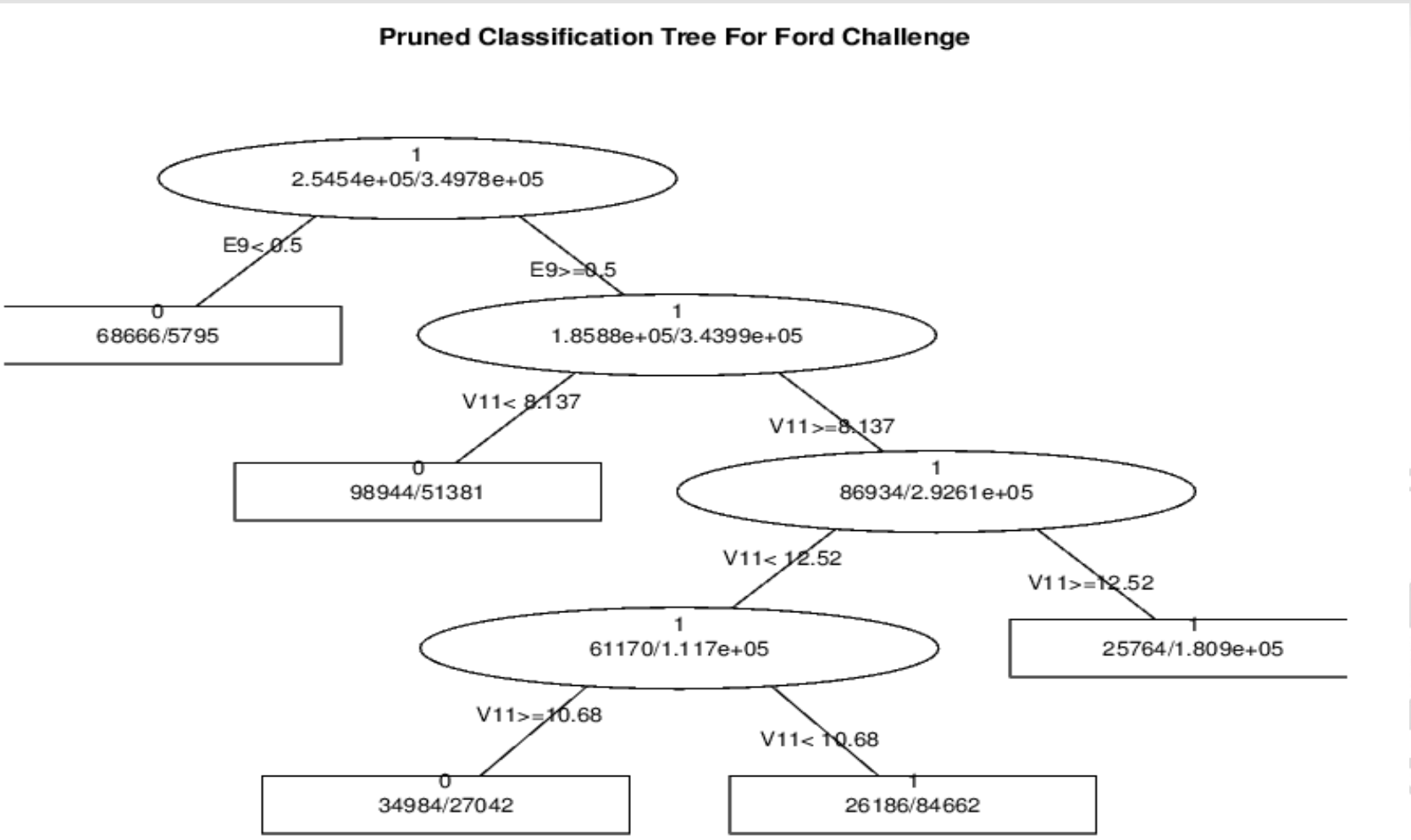
  The outcomes associated with combinations of decisions and events

- Notable observations (disadvantages)

  Instability

  Complexity

  Over fitting


Pruned Classification Tree For Ford Challenge
1
2.5454e+05/3.4978e+05
E9< 0.5
E9>=0.5
0
68666/5795
1
1.8588e+05/3.4399e+05
V11< 8.137
V11>=8.137
0
98944/51381
1
86934/2.9261e+05
V11< 12.52
V11>=12.52
1
61170/1.117e+05
1
25764/1.809e+05
V11>=10.68
V11< 10.68
0
34984/27042
1
26186/84662

Results of Decision Tree Model

Hits = 93710

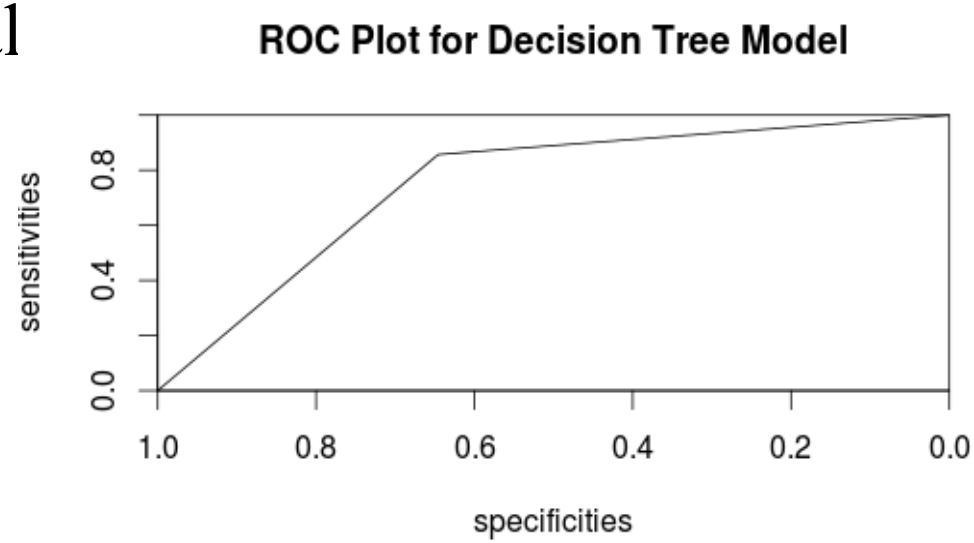
Total = 120840

Cost = 27130

Accuracy = Hits/Total

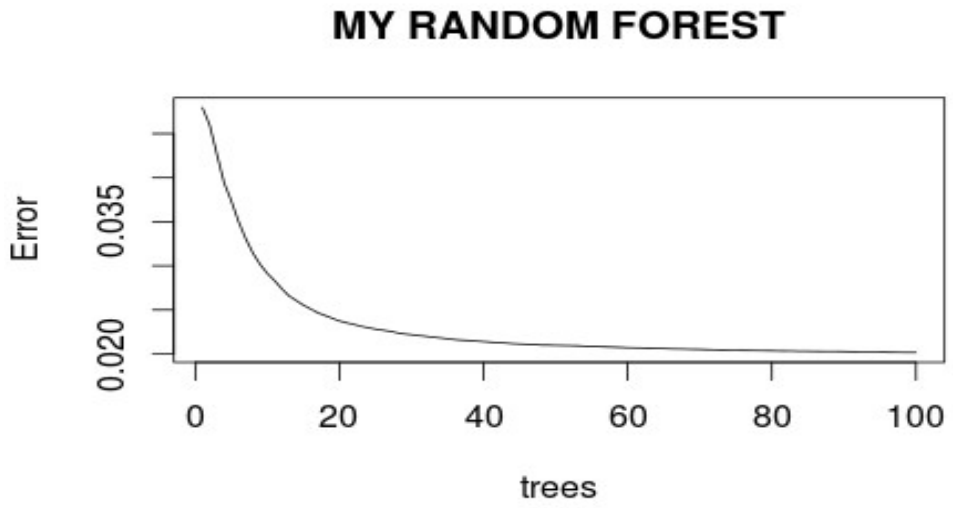
Accuracy = 77.5%

	Actual	Predicted	Count
1	0	0	18233
2	1	0	15449
3	0	1	11681
4	1	1	75477



Random Forest Model

- Random Forests is a construction of multiple decision trees which performs prediction based on majority of decisions of the individual trees.
- Number of trees to generate the forest was set to 100. This means that each row feature vector is predicted by 100 decision trees and the mode of decisions is the output of the model.
- The mean error stabilized out at 40 tree benchmark.



Results of Random Forest Model

Hits = 106583

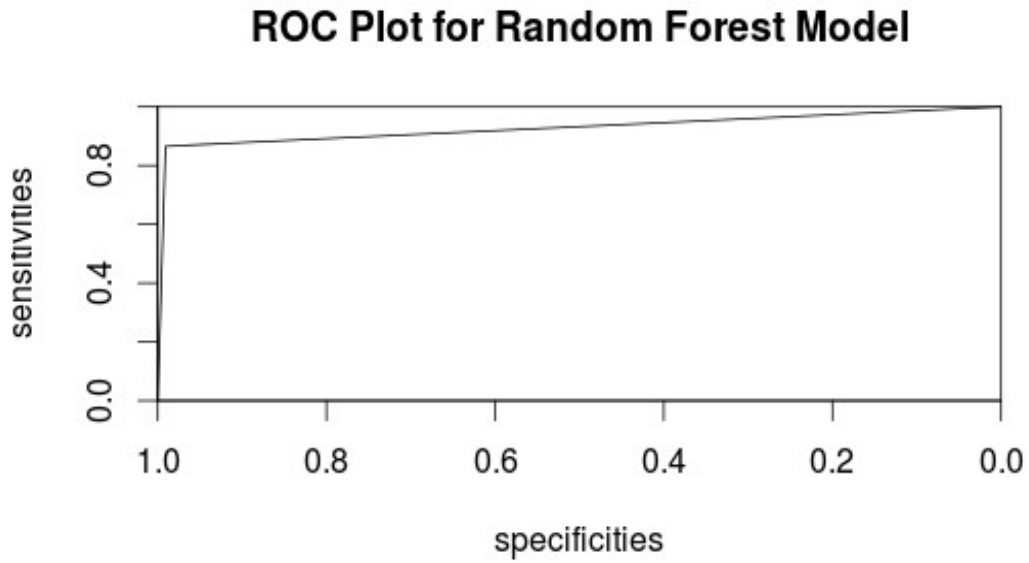
Cost = 14257

Total = 120840

Accuracy = Hits/Total

Accuracy = 88%

	Actual	Predicted	Count
1	0	0	15766
2	1	0	109
3	0	1	14148
4	1	1	90817



Questions?

Thank You!