

# Decreasing Traffic Accidents by Predicting High-risk Drivers

Bei Xia  
Yuan Gao



# Raw 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	AA	AB	AC	AD	AE	AF	AG	
A1	A12A	A12B	A12B1	A12B2A	A12B2B	A12B2C	A12B2C2	A12B3A	A12B3B	A12B3C	A12B3C2	A12C	A3	A4	A41	A5	A51	A512	B1	B11B1	B2	B2O	B3	B4	B5	B6	B6O	B7	B8A	B8B	B8C	B8D	
1													2	1	3	2			2		3		1	1	2	2		5	1	1	2	3	
1													5	1	3	2			2				1	6	9	4		1	6	2	5	5	
1													6	1	3	1	1		2				2	6	3	5		1	5	1	3	5	
1													3	1	2	2			2				1	7	2		4	7	2	3			
1													4	1	3	2			2		3		1	3	2	3		5	3	1	2	2	
1													5	1	2	1	2	1	1	19649			1	1	2	2		1	1		3		
1													6	1	3	2			2		5		1	7	3	4		1	7	1	3	6	
2	2	1	5									2																					
1													3	1	3	2			1	24733	2		1	7	3	2		5	7	2	2	4	
1													8	1	1	1	1		1	3565	2		2	8	3	5		1	7	5	2		
2	1	1	4									2																					
1													8	1	2	2			1	29822	4		1	8	6	5		1	6	3	5	7	
2			1	3									2	2		1	1		1	6280	4		1	6	5	7		5	1	1	5	8	
1	2	1	3									5																					
1													8	1	2	2			1	9601	2		3	3	3	4		4	3	2	2	2	
1														2		2			2		1		1	8	2	2		6	8	3	2	2	
1													8	1	2	2			1	1841			2	2	3	3		5	2	1	3		
1													8	1	1	2			1	17197	3		2	4	3	5		5	3	2	3	6	
1													2	1	2	2			1	36426	5		1	8	3	4		6	8	2	3	5	
1														2		1	1		1	1679			1	3	2	2		7	3	1	2	3	
1													5	1		1	1		2			4		1	1	1		4	1	1	2	2	
1													4	2		2			2			3		1	3	1	2	5	2	1	2	3	
1													8	1		2			1	4806	1		1	6	2	1		1	6	2	3	1	
1														2		2			2		2		2	7	3	3		1	7				
1													4	1	3	2			1	1680			1	5	2	2		1	5	3		2	
1													6	2		2			1	11305			1	6	2	1		1	4	2	3	2	
2	1	2		1	2	2	4					1																					
2	1	2																															
2	1	2			3	2	1					2																					
2	1	2			3	2	2	5																									
2	1	1	4									4																					
2	1	1	4									2																					
2	2	1	1									1																					
2	1	3						3	3	2	1																						
2	1	3						3	4	2	5	2																					
2	1	3										3																					

Rows(instance):10481

Columns(attribute):86

*Lots of missing value*



# Filtered Dataset

capacity	purpose	use day per week	average duration	violation	gender	age	education	job	income	city of motorcycle	accident
1	5	1	2	2	2	3	6	13	6	1	0
1	4	7	3	2	2	5	1	22	1	3	0
1	5	3	2	2	1	5	4	13	6	8	0
1	1	1	3	2	1	6	4	22	3	8	0
1	1	7	3	2	2	3	3	12	3	8	0
1	5	7	2	2	2	6	5	15	5	22	0
2	1	7	2	2	1	6	3	15	7	22	0
1	1	6	5	2	2	5	3	12	5	1	1
3	4	3	2	2	2	5	2	22	2	21	0
1	6	8	2	2	2	7	2	22	1	21	0
2	5	2	3	2	2	7	1	22	1	21	0
1	7	3	2	2	2	6	2	12	8	21	0
1	5	2	2	2	1	8	6	15	9	21	0
1	1	6	3	2	2	6	1	14	3	21	0
1	1	4	3	2	2	3	3	14	4	21	0
2	5	2	3	2	2	6	2	22	1	1	1

Rows: 1332 Columns:12

*Delete missing value*

*Attribute Reduction Method*

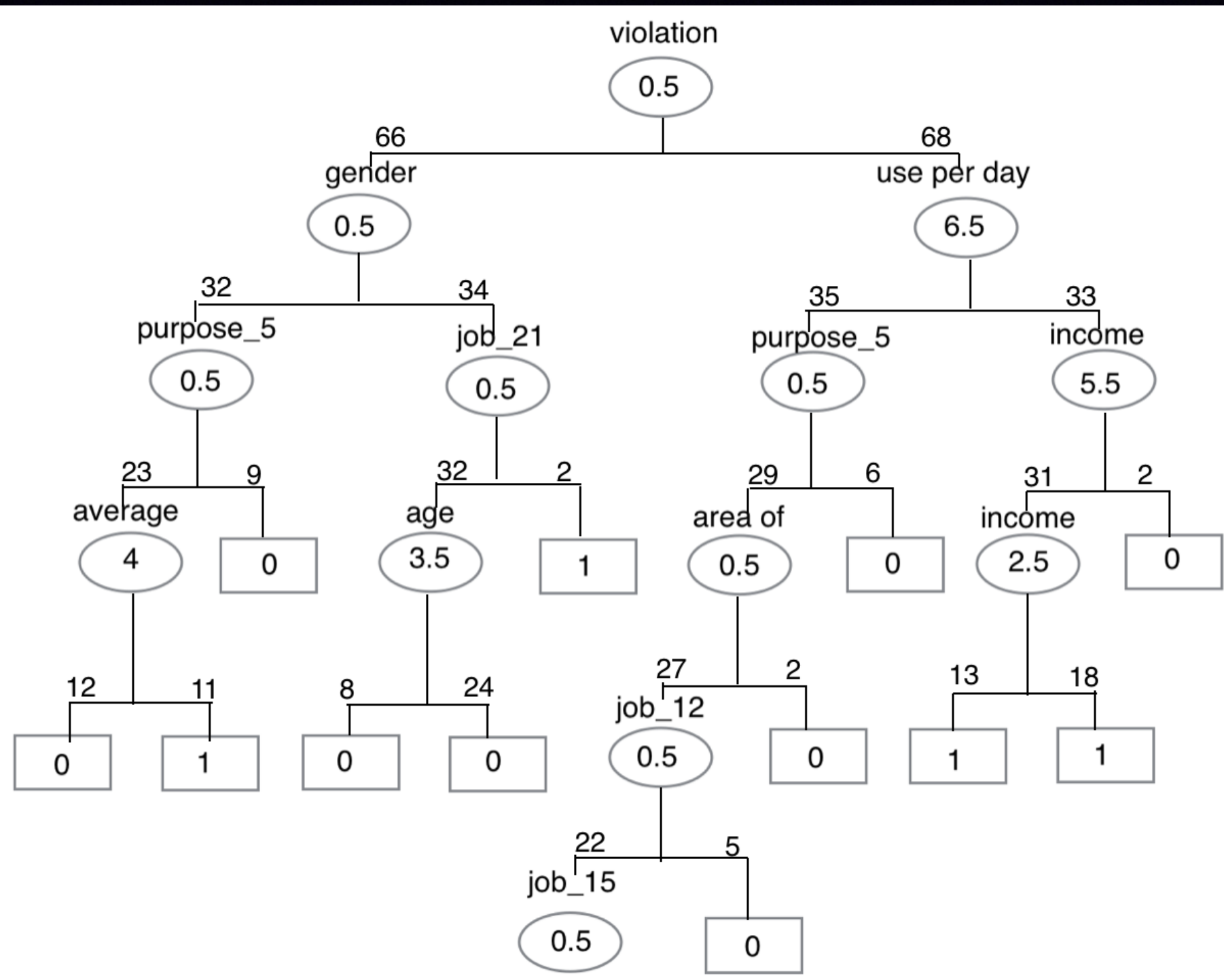


# Attribute Description

Name	Description	Type
capacity	number of passenger	Numerical
purpose	main purpose of usage	Categorical
use day per	use times per week	Numerical
average duration	average driving time	Categorical
violation	violation in a whole year	Categorical
gender	gender	Categorical
age	age	Categorical
education	education level	Categorical
job	occupation	Categorical
income	monthly salary	Categorical
city of vehicle	county where drive live	Categorical
accident	times of accidents	Categorical



# Decision Tree





# Classification

- Filter in Weka(NumericToNominal)
- Classify
  - NaiveBayes
  - Neural Network
  - KNN
  - Logistic



# Evaluation

Cross-Validation 10 Folds

Weighted Avg.	<b>TP Rate</b>	FP Rate	Precision	Recall	F-Measure
NaiveBayes	0.918	0.923	0.851	0.918	0.883
<b>Neural Network</b>	<b>0.923</b>	<b>0.923</b>	<b>0.851</b>	<b>0.923</b>	<b>0.886</b>
KNN	0.921	0.923	0.851	0.921	0.885
Logistic	0.921	0.923	0.851	0.921	0.885



# Evaluation

## Classification Confusion Matrix

	Predicted Class	
Actual Class	0	1
0	0	103
1	6	1223

NaiveBayes

	Predicted Class	
Actual Class	0	1
0	0	103
1	0	1229

Neural Network

	Predicted Class	
Actual Class	0	1
0	0	103
1	2	1227

KNN

	Predicted Class	
Actual Class	0	1
0	0	103
1	2	1227

Logistic



# Continue Work

- Hidden NaiveBayes
- And so on.



“There is no limit to knowledge.”

—Albert Einstein