



Improving Road Safety by Profiling Different Accident Type



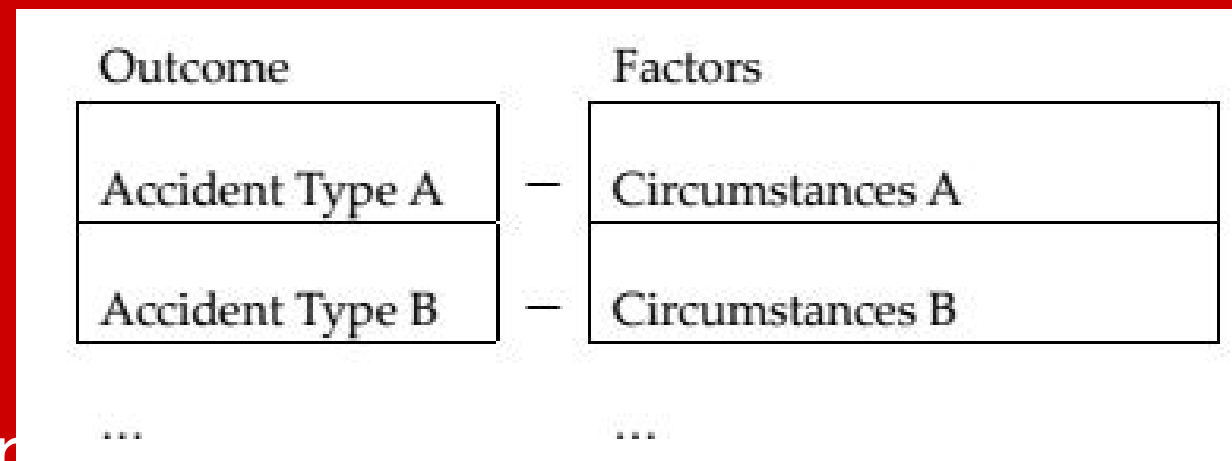
Team 7 : Angela, Aylada, Celia, Dobby, Mahsa



DATA MINING GOAL



Profile certain accident types by identifying the factors that might be the cause of that accident.



Accident types we are focusing on.

FrontHit, SideHit, BackHit, PedCrossing, and Scratch

BUSINESS GOAL



To assist Transportation Department of Taipei City Government(Client) to improve the road safety in Taipei **with more efficient way** by learning the conditions that distinguished different type of accidents.

To help the government decrease in number of certain accident types while **wisely use the budgets**.

To have a safer road usage for Taipei **citizens**.

DATA DESCRIPTION



Data set: Accident records in Taipei from [Data Taipei](#)

1	CASE_NO	YEAR	Hour_Bi	RushH	Weekda	Weatl	Light	Speed	Road	Acc_location	Paven	Signal	AccType_Bir
2	C5103031	2013	DayTime	NoRush	WeekEnd	Sunny	Light	50	Lane	MotorcycleLan	Wet	None	BackHit
3	C5103033	2013	DayTime	NoRush	WeekEnd	Sunny	Light	50	Lane	NormalLane	Wet	None	BackHit
4	C5103034	2013	DayTime	NoRush	WeekEnd	Sunny	Light	50	Lane	FastLane	Wet	None	Other
5	C5103035	2013	DayTime	RushHo	WeekDay	Sunny	Light	50	Lane	MotorcycleLan	Wet	None	Scratch
6	C5103037	2013	DayTime	NoRush	WeekDay	Sunny	Light	50	Fork	CrossRoad	Wet	TrafficPe	Other
7	C5103038	2013	DayTime	NoRush	WeekDay	Cloudy	Light	40	Fork	CrossRoad	Wet	TrafficPe	Scratch
8	C5103061	2013	DayTime	NoRush	WeekDay	Sunny	Light	40	Lane	NormalLane	Wet	None	Other
9	C5103063	2013	NightTime	NoRush	WeekEnd	Sunny	Light	50	Lane	SlowLane	Wet	None	BackHit
10	C5103068	2013	DayTime	NoRush	WeekDay	Sunny	Light	50	Fork	CrossRoad	Wet	TrafficSi	SideHit

Time variables

Environmental variables

DATA DESCRIPTION (Cont.)



2013 - 39,577 rows 30 columns → 17,991 rows 11 columns
2012 - 39,062 rows 30 columns → 17,843 rows 11 columns
2011 - 41,082 rows 30 columns → 18,963 rows 11 columns

We combined the records of same case ID
one row: one record => one row: **one accident case**

We also deleted records that has accident on highway. (not our goal)

To reduce dimensionalities, we have

- deleted some columns that are not significant in profiling accident types
- binned the categories that are similar

DATA DESCRIPTION (Cont.)



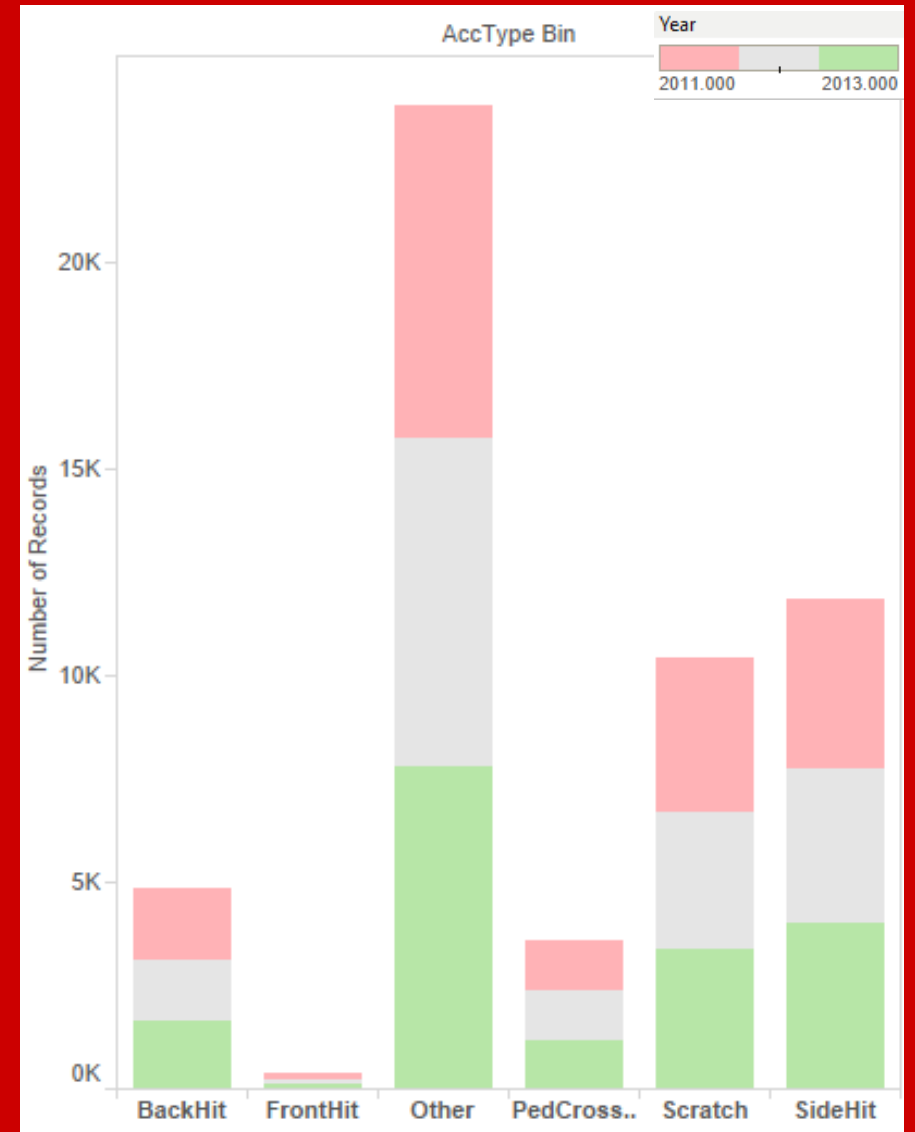
We found that our data is

imbalanced

To deal with this,
we need to

oversample

our data.



METHODS

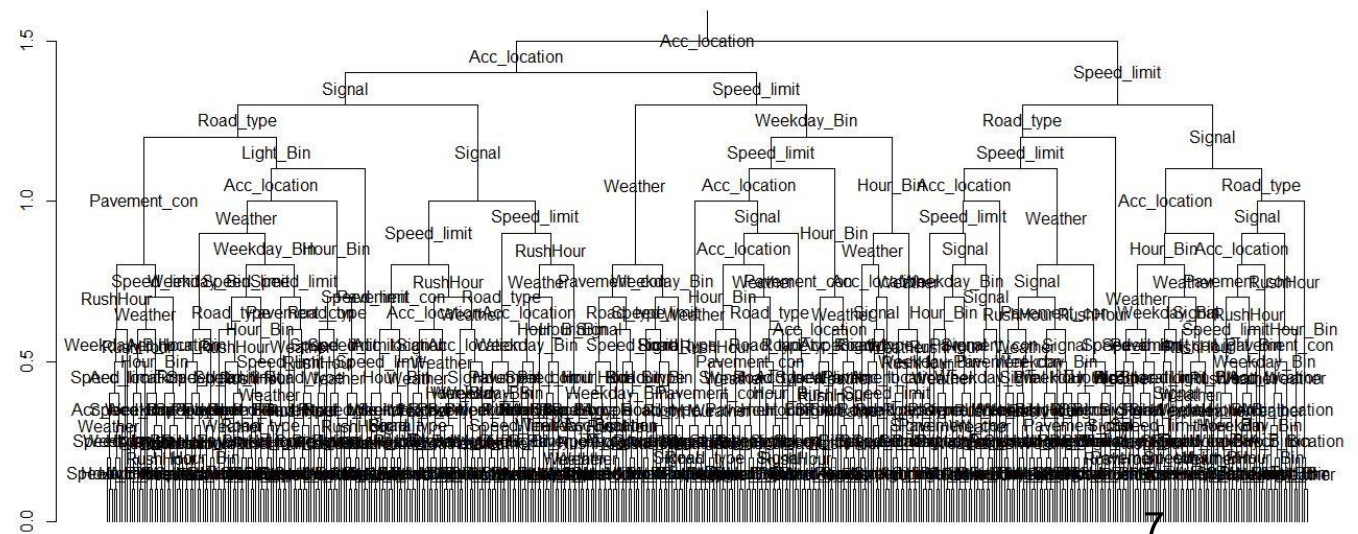
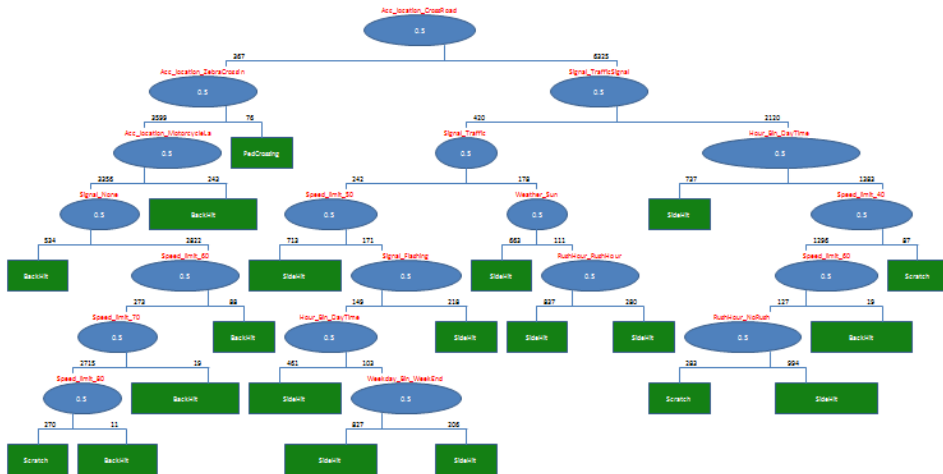
Discriminant Analysis (R)

Decision Tree (R)

Decision Tree (XLminer)

Random Forest (R)

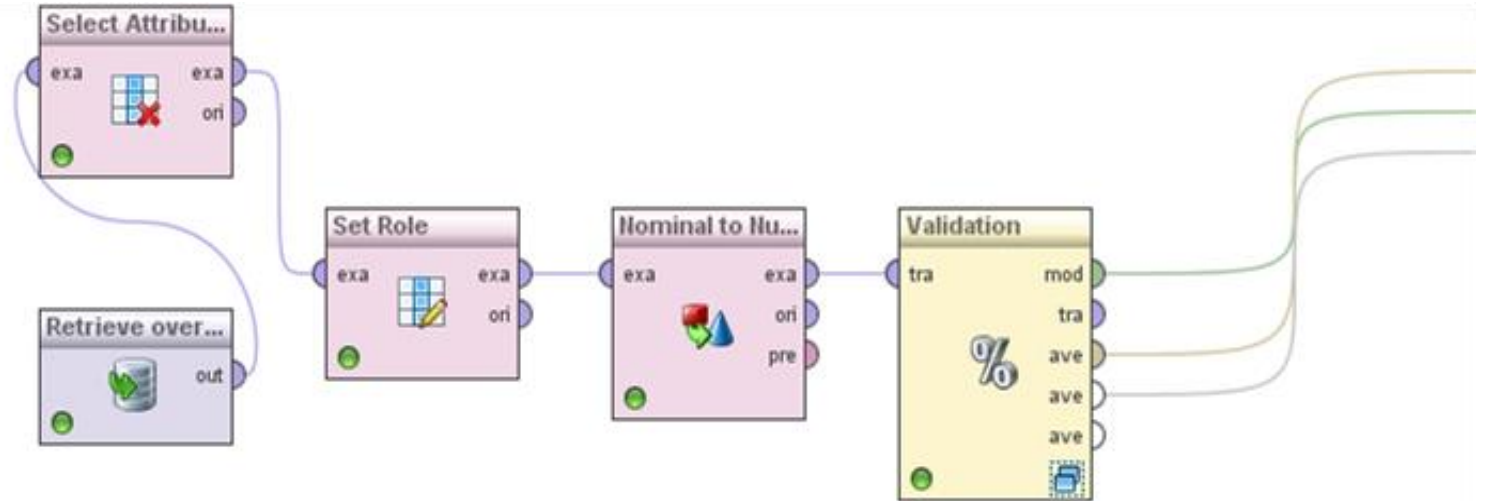
Decision Tree (RapidMiner)



Method Chosen-Decision Tree (RapidMiner)

Main process

Data Preparation
(and Oversampling)



Decision Tree

criterion: gain_ratio

maximal depth: 7

apply pruning

confidence: 0.25

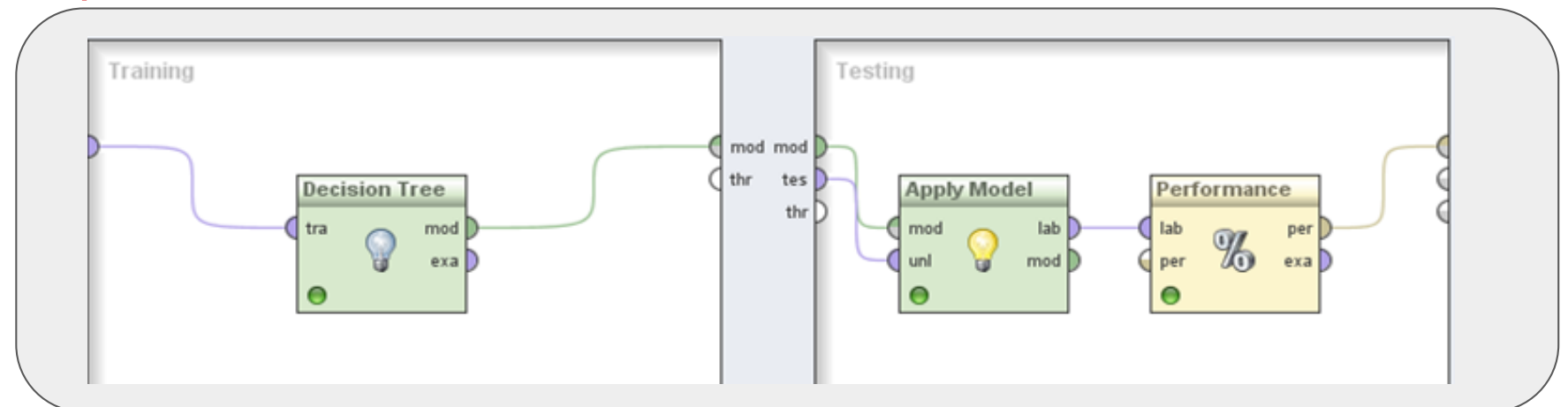
apply prepruning

minimal gain: 0.06

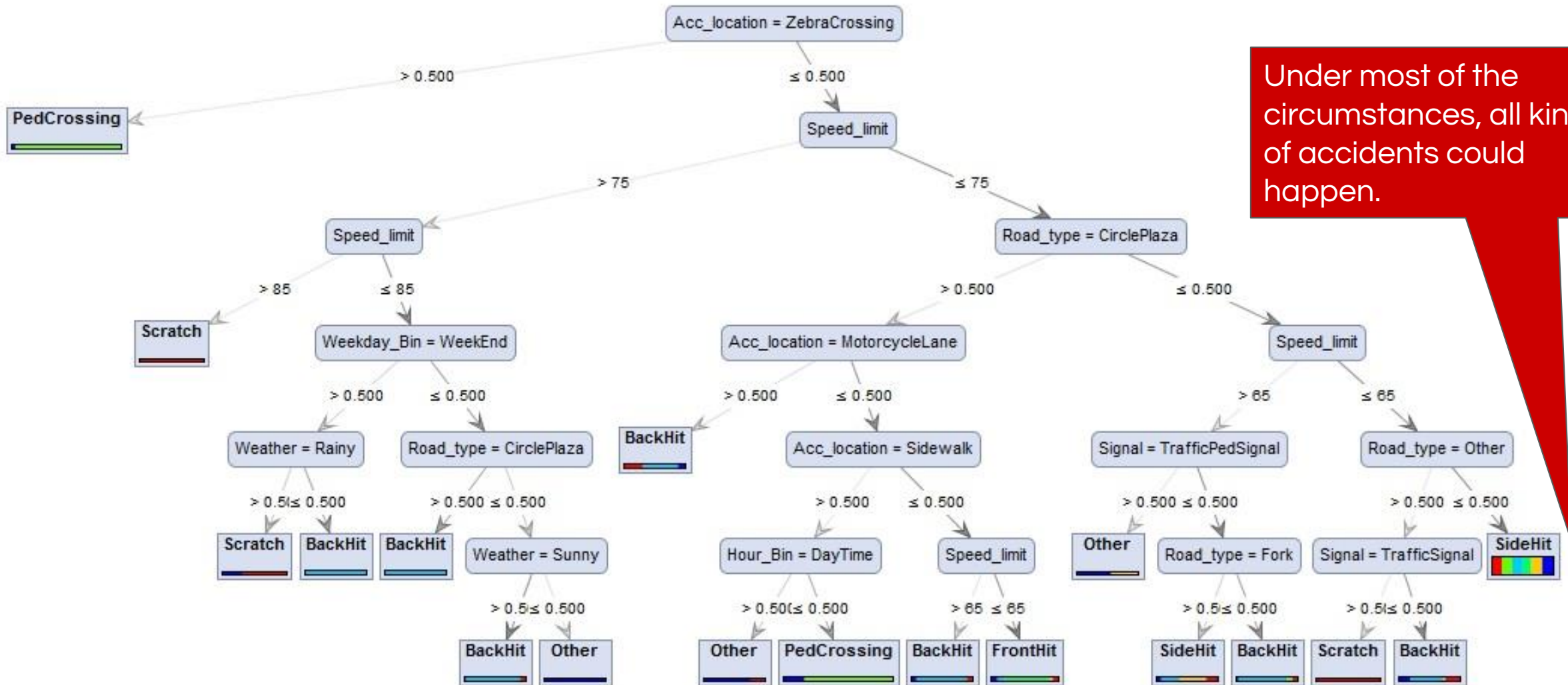
minimal leaf size: 2

minimal size for split: 10

Sub-process



Method Chosen - Decision Tree (RapidMiner)



Under most of the circumstances, all kinds of accidents could happen.

Results

However, there are circumstances where only certain accident type happens.

Outcome	Factors
FrontHit	1. At Circle/Plaza, Speed limit ≤ 65 , Fast/Slow/Normal Lane
BackHit	1. Sunny, Speed limit > 65 , Non-Fork, Non-Cross Road 2. Other Road Type, Speed limit < 65 , Non-Traffic Signal 3. Other Road Type, Speed limit < 65 , Motorcycle Lane 4. At Circle/Plaza, Motorcycle Lane
SideHit	1. Speed limit 65~75, when pavement is Wet, at Cross Road
PedCrossing	1. At Zebra Crossing 2. At Circle/Plaza, Sidewalk, Night Time, Speed limit ≤ 75
Scratch	1. Speed limit > 85 2. Speed limit ≤ 85 , Rainy, Weekend

	Factors
1	Accident Location
2	Speed Limit
3	Road Type
4	Signal
5	Weather
6	Weekday/Weekend

Performance Evaluation

1. Cross Validation (with 10 Validations)

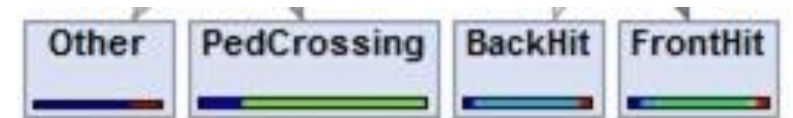
Those cannot be distinguished very well are predicted as SideHit

Overall Accuracy
Class of Interest

	true Other	true BackHit	true FrontHit	true PedCrossing	true SideHit	true Scratch	class precision
pred. Other	10	1	0	4	4	4	43.48%
pred. BackHit	210	1189	165	32	54	478	55.87%
pred. FrontHit	799	833	4940	224	248	586	64.74%
pred. PedCrossing	75	0	0	1515	3	1	95.04%
pred. SideHit	18903	17975	14895	18225	19691	18925	18.13%
pred. Scratch	3	2	0	0	0	6	54.55%
class recall	0.05%	5.95%	24.70%	7.58%	98.45%	0.03%	

1. Purity of the end nodes

We did identify some accident types that happens under certain circumstances



RECOMMENDATION



1. Build a **bridge** or **underground walkway** for pedestrian at the circles/ plaza to increase their safety.
2. Install more signals at the circles/ plaza to avoid front hit.
3. Install anti-slip textures and decrease the speed limit before the cross roads to decrease the number of side hit and scratch.
4. Dynamic speed limit to avoid backhit. (low speed limit in sunny days ?)

An aerial photograph of a city street, likely in New York City, showing several yellow taxis and a white van. The street is lined with buildings, and the overall scene is captured from a high angle. A large green rounded rectangle is overlaid on the center of the image, containing the text.

SAFETY IS OUR BUSINESS

SmartSign.com • 800-952-1457 • S-4137