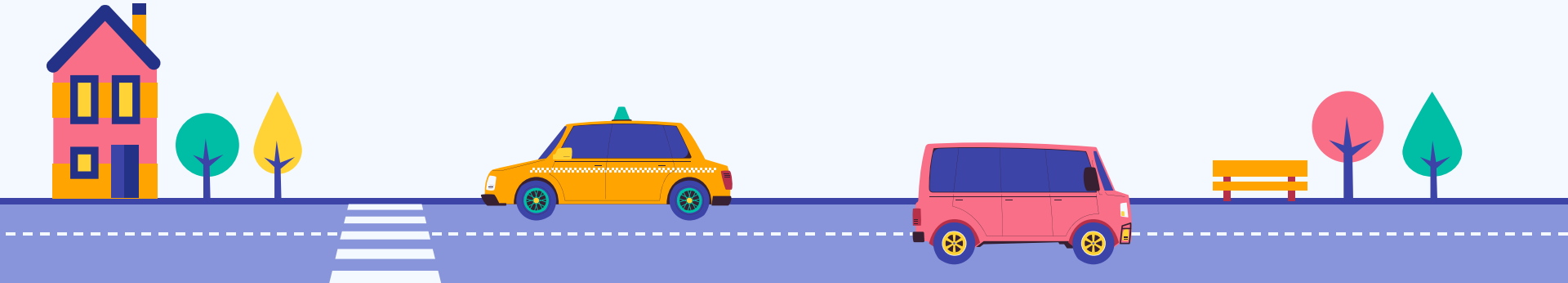
Predicting Driver At Fault In A Vehicle Collision

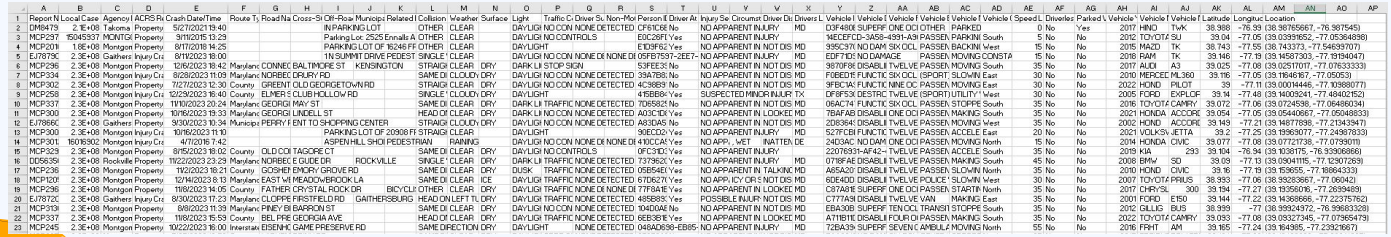
Julia Chen and Yuan Yin



Dataset

https://catalog.data.gov/dataset/crash-reporting-drivers-data

- Traffic Collisions within Montgomery County in Maryland
- Collected by Automated Crash Reporting System (ACRS) of the Maryland State Police
- 39 attributes and 184,897 instances of vehicle crashes



	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	AH	AI	AJ	AK	AL	AM	AN	AO	AP
1	Report N	Local Case	Agency I	ACRS Re	Crash Date/Time	Route Ty	Road Na	Cross-St	Off-Road	Municipa	Related I	Collision	Weather	Surface	Light	Traffic C	Driver Su	Non-Mot	Person ID	Driver Ar	Injury Se	Circumst	Driver Di	Driver L	Vehicle I	Vehicle I	Vehicle I	Vehicle I	Vehicle I	Vehicle (	Speed C	Driverle	Parked V	Vehicle '	Vehicle I	Vehicle I	Latitude	Longitud	Location			
2	DM8479	2.1E+08	Takoma	Property	5/27/2021 19:40				IN PARKING LOT			OTHER	CLEAR		DAYLIGHT	NO CONT	NONE DETECTED		CF61CB8	No	NO APPARENT	INJURY	MD		03F480E	SUPERF	ONE OCI	OTHER	PARKED		0	No	Yes	2017	HINO	TWK	38.988	-76.99	(38.98765667, -76.987545)			
3	MCP237	15045937	MONTGO	Property	3/11/2015 13:29				Parking Lot	2525 Ennalls A		OTHER	CLEAR		DAYLIGHT	NO CONTROLS			E0C28F1	Yes	NO APPARENT	INJURY			14ECEFCD	SA58-4391-A9	PASSEN	PARKIN	South		5	No	No	2012	TOYOTA	SU	39.04	-77.05	(39.03991652, -77.05364836)			
4	MCP201	1.9E+08	Montgon	Property	9/17/2016 14:25				PARKING LOT OF	16246 FF		OTHER	CLEAR		DAYLIGHT				E1D9F5C	Yes	NO APPARENT	IN NOT DIS	MD		995C37C	NO DAM	SIX OCL	PASSEN	BACKIN	West	15	No	No	2015	MAZD	TK	38.743	-77.55	(38.7433733, -77.54839707)			
5	EJ78796	2.3E+08	Gaithers	Injury Cra	8/11/2023 18:00				TN SUMMIT DRIVE	PEDEST		SINGLE '	CLEAR		DAYLIGHT	NO CONT	NONE DE	NONE DE	05FB7597-2EE7-		NO APPARENT	INJURY	MD		EDF71DE	NO DAMAGE		PASSEN	MOVING	CONSTA	15	No	No	2018	RAM		39.146	-77.19	(39.14597303, -77.19140417)			
6	MCP236	2.3E+08	Montgon	Injury Cra	12/6/2023 18:42	Maryland	CONNEC	BALTIMORE ST	KENSINGTON			STRAIGH	CLEAR	DRY	DARK LI	STOP SIGN			53FEE3C	No	NO APPARENT	IN NOT DIS	MD		9870F9E	DISABLE	TWELVE	PASSEN	MOVING	South	35	No	No	2017	AUDI	A3	39.025	-77.08	(39.02517017, -77.07633333)			
7	MCP334	2.3E+08	Montgon	Injury Cra	9/28/2023 11:09	Maryland	NORBEC	DRURY RD				SAME DI	CLOUDY	DRY	DAYLIGHT	NO CONT	NONE DE		39A7B8(	No	NO APPARENT	IN NOT DIS	MD		F0BED1E	FUNCTIC	SIX OCL	(SPORT	SLOWIN	East	30	No	No	2010	MERCEC	ML350	39.116	-77.05	(39.11646167, -77.05053			
8	MCP302	2.3E+08	Montgon	Property	7/27/2023 12:30	County	GREENT	OLD GEORGETOWN RD				STRAIGH	CLEAR	DRY	DAYLIGHT	NO CONT	NONE DETECTED		4C38689	No	NO APPARENT	IN NOT DIS	MD		9FBC1A(	FUNCTIC	NINE OC	PASSEN	MOVING	East	30	No	No	2022	HOND	PILOT	39	-77.11	(39.00014446, -77.10989077)			
9	MCP258	2.3E+08	Montgon	Injury Cra	12/29/2023 16:40	County	ELMER S	CLUB HOLLOW RD				SINGLE '	CLOUDY	DRY	DAYLIGHT				41BB68(	Yes	SUSPECTED	MINOR IN	UR	TX	DF8F53(	DESTRC	TWELVE	(SPORT)	UTILITY '	West	30	No	No	2005	FORD	EXPLOR	39.14	-77.48	(39.14003241, -77.48402152)			
10	MCP337	2.3E+08	Montgon	Property	11/19/2023 20:24	Maryland	GEORGI	MAY ST				SAME DI	CLEAR	DRY	DARK LI	TRAFFIC	NONE DETECTED		7D6582(	No	NO APPARENT	IN NOT DIS	MD		06AC74	FUNCTIC	SIX OCL	PASSEN	STOPPE	South	35	No	No	2016	TOYOTA	CAMRY	39.072	-77.06	(39.0724598, -77.06486034)			
11	MCP300	2.3E+08	Montgon	Property	10/16/2023 19:33	Maryland	GEORGI	LINDELL ST				HEAD ON	CLEAR	DRY	DARK LI	NO CONT	NONE DETECTED		A03CIDE(	Yes	NO APPARENT	IN LOCKED	MD		7BAFAB	DISABLE	ONE OCI	PASSEN	MAKING	South	35	No	No	2021	HONDA	ACCORE	39.054	-77.05	(39.05440667, -77.05048833)			
12	EJ78666	2.3E+08	Gaithers	Property	9/30/2023 10:34	Municipa	PERRY F	ENT TO	SHOPPING CENTER			STRAIGH	CLOUDY	DRY	DAYLIGHT	NO CONT	NONE DETECTED		A83DA5	No	NO APPARENT	IN NOT DIS	MD		2D8364(	DISABLE	TWELVE	PASSEN	MOVING	West	35	No	No	2002	HOND	ACCORE	39.149	-77.21	(39.14877898, -77.21343947)			
13	MCP300	2.3E+08	Montgon	Injury Cra	10/16/2023 11:10				PARKING LOT OF	20908 FI		STRAIGH	CLEAR		DAYLIGHT				90EC02(	Yes	NO APPARENT	INJURY	MD		527FCB(	FUNCTIC	TWELVE	PASSEN	ACCELE	East	20	No	No	2021	VOLKSV	JETTA	39.2	-77.25	(39.19963077, -77.24987833)			
14	MCP301	1601839(	Montgon	Injury Cra	4/7/2016 7:42				ASPEN HILL	SHOPPEDESTRIAN			RAINING		DAYLIGHT	NO CONT	NONE DE	NONE DE	410CCAE	Yes	NO APPA	WET	INATTEN	DE	24D3AC	NO DAM	ONE OCI	PASSEN	MOVING	North	15	No	No	2014	HONDA	CIVIC	39.077	-77.08	(39.07721738, -77.0759011			
15	MCP329	2.3E+08	Montgon	Property	8/15/2023 18:02	County	OLD COI	TAGORE CT				SAME DI	CLEAR	DRY	DAYLIGHT	NO CONT	ROLS		0FC31D(	Yes	NO APPARENT	INJURY			2207693	AF42-	TWELVE	PASSEN	ACCELE	South	35	No	No	2019	KIA	293	39.104	-76.94	(39.1039175, -76.93906866)			
16	DD56359	2.3E+08	Rockville	Property	11/22/2023 23:29	Maryland	NORBEC	E GUDE DR		ROCKVILLE		SINGLE '	CLEAR	DRY	DARK LI	TRAFFIC	NONE DETECTED		73736C(	Yes	NO APPARENT	INJURY	MD		0718FAE	DISABLE	TWELVE	PASSEN	MAKING	South	45	No	No	2008	BMW	5D	39.09	-77.13	(39.09041115, -77.12907269)			
17	MCP236	2.3E+08	Montgon	Property	11/2/2023 18:21	County	GOSHEN	EMORY GROVE RD				SAME DI	CLEAR	DRY	DUSK	TRAFFIC	NONE DETECTED		05B54E(	Yes	NO APPARENT	IN TALKING	MD		A65A20	DISABLE	TWELVE	PASSEN	SLOWIN	North	35	No	No	2010	HOND	CIVIC	39.16	-77.19	(39.1598655, -77.18964333)			
18	MCP120	2.3E+08	Montgon	Property	12/14/2023 8:13	Maryland	EAST WI	MEADOWBROOK LA				SAME DI	CLEAR	ICE	DAYLIGHT	TRAFFIC	NONE DETECTED		67D627(	Yes	NO APP, ICY	OR S	NOT DIS	MD	6DE40D	DISABLE	TWELVE	POLICE '	SLOWIN	West	30	No	No	2007	TOYOTA	PRIUS	38.993	-77.06	(38.9928667, -77.06042)			
19	MCP296	2.3E+08	Montgon	Property	11/6/2023 14:05	County	FATHER	CRYSTAL ROCK DR		BICYCLI		OTHER	CLEAR	DRY	DAYLIGHT	TRAFFIC	NONE DETECTED	NONE DE	77F9A1E	Yes	NO APPARENT	IN LOCKED	MD		C87A81E	SUPERF	ONE OCI	PASSEN	STARTIN	North	35	No	No	2017	CHRYSL	300	39.194	-77.27	(39.1935696, -77.2699489)			
20	EJ78720	2.3E+08	Gaithers	Injury Cra	8/30/2023 17:23	Maryland	CLOPPE	FIRSTFIELD RD		GAITHERSBURG		HEAD ON	LEFT TU	DRY	DAYLIGHT	TRAFFIC	NONE DETECTED		485B93(	Yes	POSSIBLE	INJUR	NOT DIS	MD	C777A9(	DISABLE	TWELVE	VAN	MAKING	East	35	No	No	2001	FORD	E150	39.144	-77.22	(39.14368666, -77.2275752)			
21	MCP313	2.3E+08	Montgon	Property	8/8/2023 11:39	Maryland	PINEY BI	BARRON ST				SAME DI	CLEAR	DRY	DAYLIGHT	NO CONT	NONE DETECTED		1040DAE	No	NO APPARENT	IN NOT DIS	MD		EBA30B	SUPERF	TEN OCL	TRANSIT	STOPPE	South	35	No	No	2012	GILLIG	BUS	38.999	-77	(38.99924772, -76.99683328)			
22	MCP337	2.3E+08	Montgon	Property	11/8/2023 15:59	County	BEL PRE	GEORGIA AVE				HEAD ON	CLEAR	DRY	DAYLIGHT	TRAFFIC	NONE DETECTED		6EB3B1E	Yes	NO APPARENT	IN LOCKED	MD		A71B1E	DISABLE	FOUR OI	PASSEN	MAKING	South	35	No	No	2022	TOYOTA	CAMRY	39.093	-77.08	(39.09327345, -77.07965479)			
23	MCP245	2.3E+08	Montgon	Property	10/22/2023 16:00	Interstate	EISENHC	GAME PRESERVE RD				SAME DI	RECTION	DRY	DAYLIGHT		NONE DETECTED		046AD38-EB85-		NO APPARENT	INJURY	MD		72BA39(	SUPERF	SEVEN C	AMBULA	MOVING	North	55	No	No	2016	FRHT	AM	39.165	-77.24	(39.164365, -77.23921667)			

Goal

- Predict the Driver at Fault class
- Settle disagreements in a vehicle collision

Class Distribution:

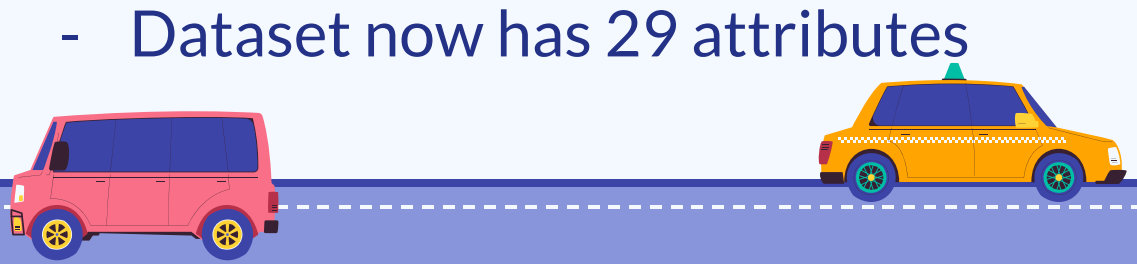
- 52% Yes
- 45% No

Initial Problems

- Issues opening in Weka
  - Apostrophes
  - New line in values
  - Double quotes

Preprocessing

- Remove Instances Missing the Class value
  - 184,897 instances → 180,211 instances
- Remove Attributes Missing > 70% of Their Values
- Remove Derived Attributes
  - Location vs Longitude/Latitude
- Remove Unnecessary Attributes
  - Driverless Vehicle, Person ID, Vehicle ID
  - Dataset now has 29 attributes



Preprocessing

- Stratified Random Sampling
  - 180,211 instances → 10,012 instances
- Unify Values
  - “CLEAR” vs. “Clear” and “Montgomery” vs. “Montgomery Police Department”
- Fill in Missing Values
  - Mean & Mode

Train-Validation-Test Split

80%

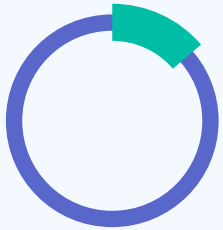
Train

8,008 instances

10%

Validation

1,002 instances



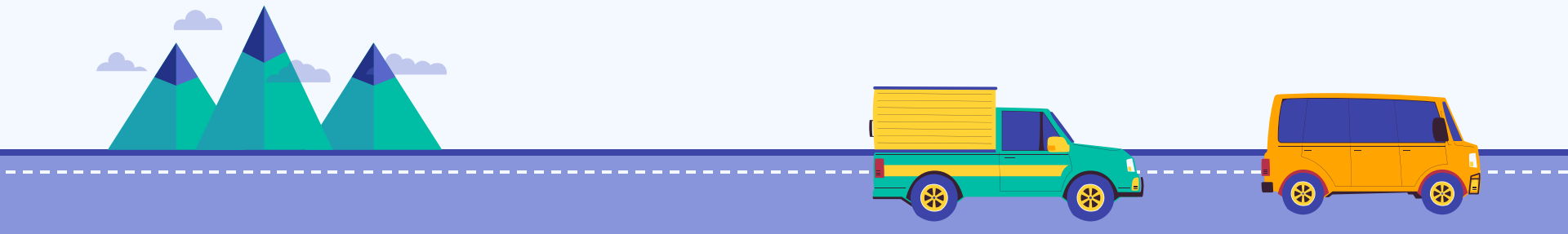
10%

Test

1,002 instances

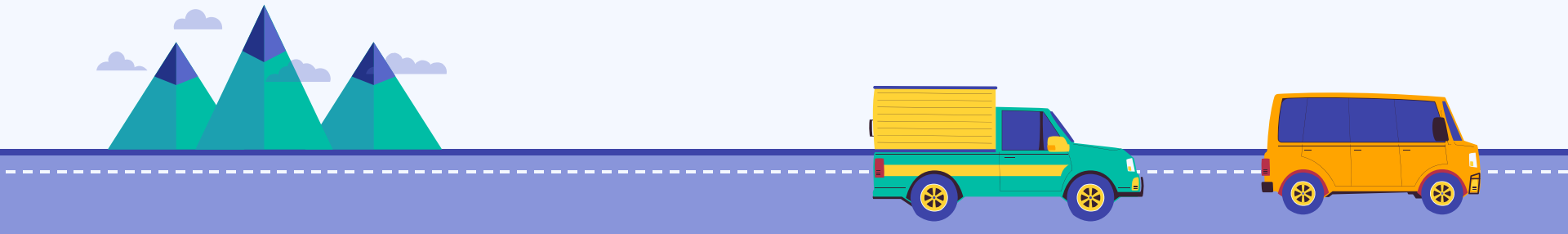
Attribute Selection

- CorrelationAttributeEval (11 attributes)
  - Evaluates the correlation between an attribute and the class variable
- CfsSubsetEval (3 attributes)
  - Selects attributes that have a high predictive ability and low intercorrelation



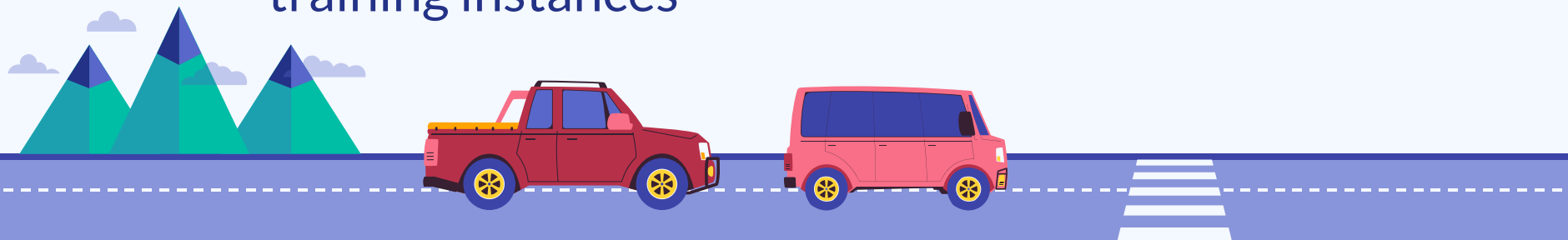
Attribute Selection

- OneRAttibuteEval (13 attributes)
  - Evaluates attributes using OneR classifier
- InfoGainAttributeEval (9 attributes)
  - Selects attributes based on how much information is gained from that attribute with respect to the class
- Intuition (18 attributes)



Classifiers

- NaïveBayes
  - Assume attributes are independent
  - Makes predictions by combining the likelihood of individual attributes
- KStar
  - Memorizing training instances
  - Comparing testing instance to previously existing training instances



Classifiers

- DecisionTable
  - Build and use a simple decision table using majority combinations
- OneR
  - Creates a ruleset based on one attribute

CorrelationAttributeEval

	Accuracy	True Positive	False Positive	ROC Area
NaïveBayes	76.3473	0.700	0.163	0.876
KStar	80.8383	0.828	0.215	0.883
DecisionTable	80.4391	0.834	0.230	0.886
OneR	69.5609	0.879	0.515	0.682

CfsSubsetEval

	Accuracy	True Positive	False Positive	ROC Area
NaïveBayes	74.1517	0.616	0.114	0.823
KStar	74.1517	0.601	0.097	0.824
DecisionTable	74.5509	0.612	0.101	0.829
OneR	45.01	0.000	0.032	0.484

OneRAttributeEval

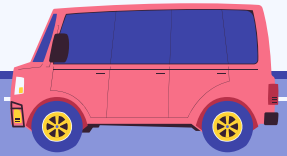
	Accuracy	True Positive	False Positive	ROC Area
NaïveBayes	78.6427	0.881	0.322	0.881
KStar	69.5609	0.879	0.515	0.682
DecisionTable	80.1397	0.802	0.200	0.882
OneR	69.5609	0.879	0.515	0.682

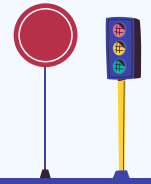
InfoGainAttributeEval

	Accuracy	True Positive	False Positive	ROC Area
NaïveBayes	73.5529	0.759	0.292	0.836
KStar	74.3513	0.769	0.285	0.797
DecisionTable	79.5409	0.866	0.285	0.881
OneR	45.01	0.000	0.032	0.484

Intuition

	Accuracy	True Positive	False Positive	ROC Area
NaïveBayes	77.5449	0.817	0.273	0.867
KStar	75.1497	0.748	0.245	0.824
DecisionTable	80.1397	0.802	0.200	0.886
OneR	69.5609	0.879	0.515	0.682



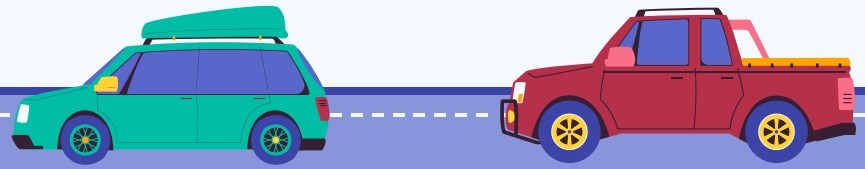


Best Model

DecisionTable classifier with CorrelationAttributeEval

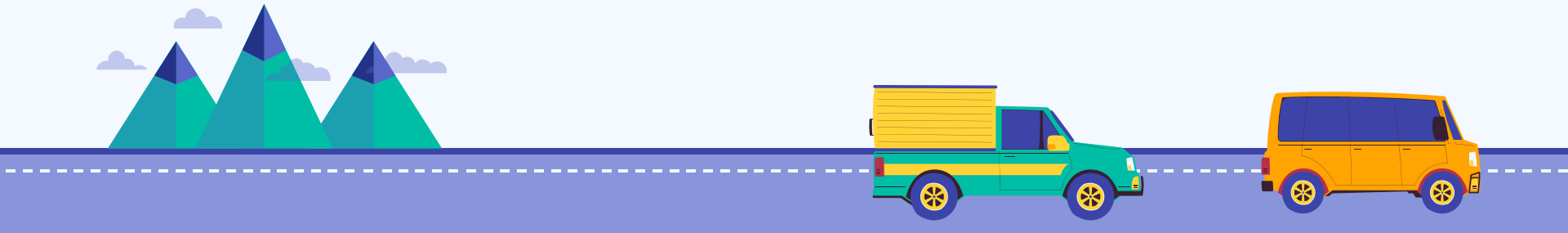
Accuracy	True Positive	False Positive	ROC Area
80.4391	0.834	0.230	0.886

- Second Highest Accuracy
- One of the highest True Positive Rates
- Highest ROC Area



Potential Improvements

- DecisionTable classifier with CorrelationAttributeEval could be improved by optimizing the cutoff value for correlation analysis



Thank you!