# AUTO INSURANCE FRAUD DETECTION

Rampam Greeshma Geethika 20BCD7094

#### INTRODUCTION

- AUTO INSURANCE FRAUD IS A SERIOUS ISSUE THAT COSTS INDUSTRIES MILLIONS OF DOLLARS ANNUALLY.
- DUE TO THE COMPLEXITY OF IDENTIFYING FRAUDULENT CLAIMS AND THE LARGE AMOUNTS OF DATA INVOLVED,
- DETECTING AND PREVENTING FRAUD IS DIFFICULT. INSURERS CAN NOW QUICKLY AND PRECISELY IDENTIFY FRAUDULENT CLAIMS THANKS TO MACHINE LEARNING,
- A PROMISING SOLUTION FOR AUTO INSURANCE FRAUD DETECTION.

## RELATED WORKS

1. AUTOMOBILE INSURANCE FRAUD DETECTION USING SUPERVISED CLASSIFIERS

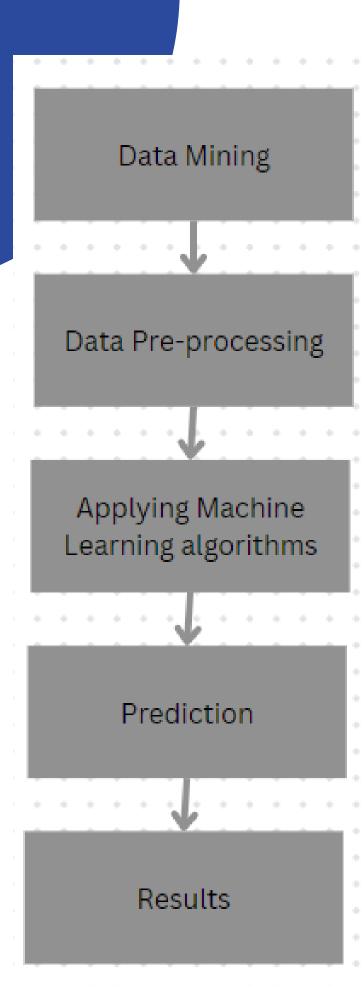
80.4%

2. AUTO INSURANCE FRAUD
DETECTION WITH MULTIMODAL 81.81%
LEARNING

3. AUTO INSURANCE CLAIM FRAUD ANALYTICS

75%





# METHODOLOGY

## ALGORITHMS USED

#### **Boosting Algorithms**

- Light GBM
- Gradient Boosting Algorithm
- CAT Boosting Algorithm

#### **Machine Learning Algorithms**

- Decision Tree Classifier
- SVM Algorithm
- Random Forest Classifier
- KNN Algorithm

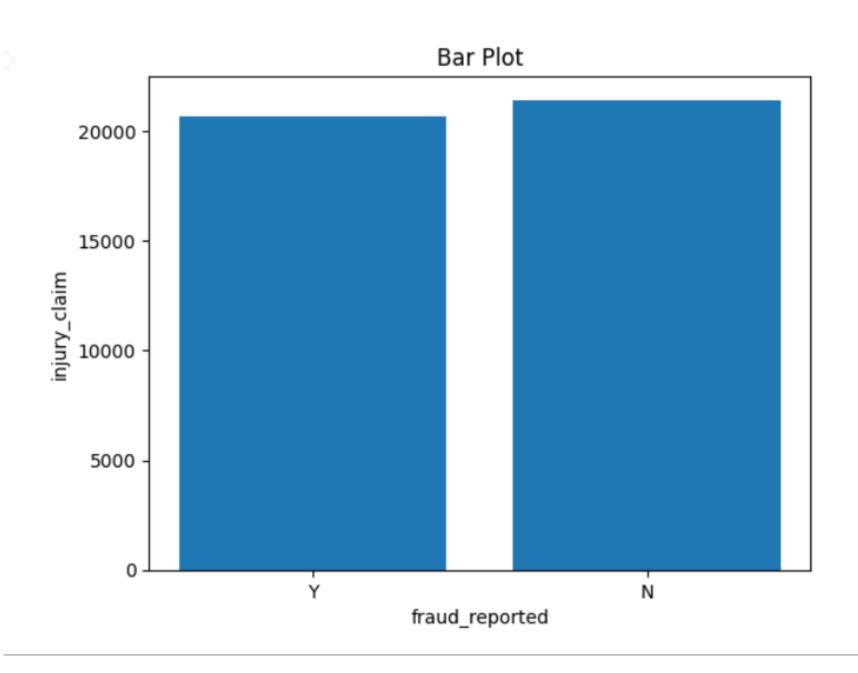
- ADA Boosting Algorithm
- XG Boost Algorithm

#### **Data Collection**

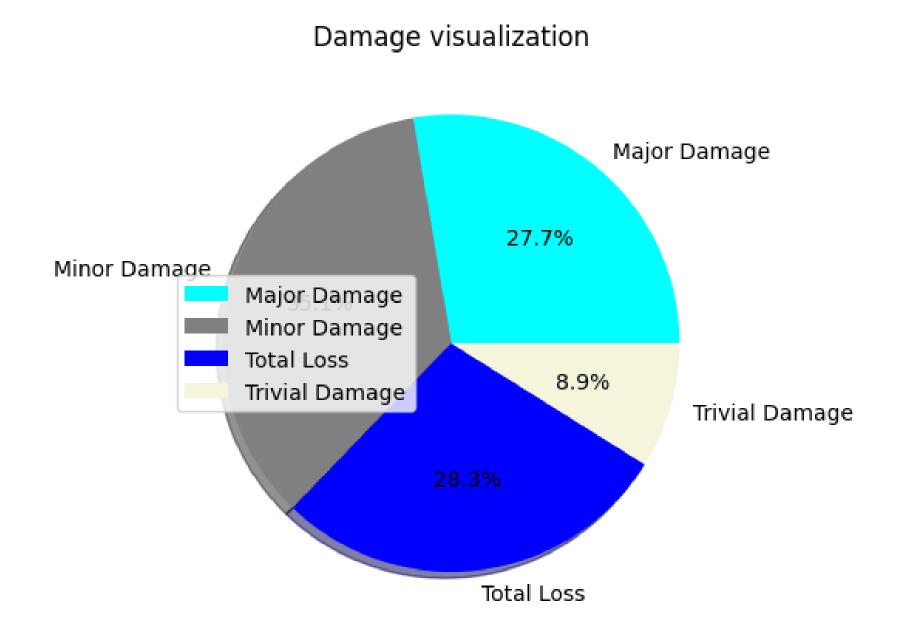
- Hybrid Datasets from Kaggle
- 1102 Instances
- Target Column: fraud\_reported

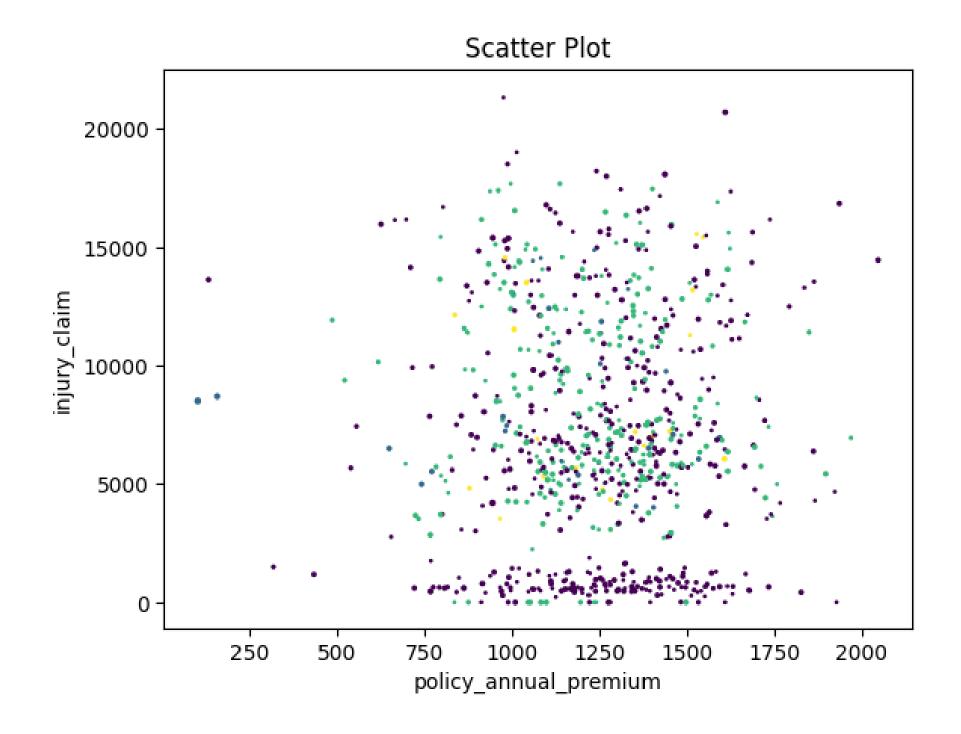
police_report_available	total_claim_amount	injury_claim	property_claim	vehicle_claim	auto_make	auto_model	auto_year	fraud_reported
YES	71610	6510	13020	52080	Saab	92x	2004	Υ
YES	20000	8512	5263	55000	Saab	82x	2005	Υ
YES	51590	9380	9380	85274	Audi	A5	2015	N
YES	27700	2770	2770	22160	Toyota	Camry	2012	N
NO	42300	6452	4700	54875	Saab	92x	1996	N

## Data Visualization



months_as_customer	1	0.87	0.0045	0.0026	0.061	0.015	0.039	-0.0022	0.027	0.057	0.0054	-0.0015	0.045	0.008	0.058	-0.0096	0.038	-0.011
age ·	0.87	1	0.067	0.051	0.0016	0.029	0.026	-0.014	-0.003	0.08	0.014	-0.0011	0.07	0.077	0.067	0.073	0.055	0.0071
policy_number	0.0045	0.067	1	0.0029	-0.043	0.014	-0.02	0.02	-0.013	0.088	0.034	0.03	-0.0075	0.006	-0.0019	0.0097	0.049	0.041
policy_deductable -	0.0026	0.051	0.0029	1	-0.074	0.0057	-0.041	0.029	0.0032	0.072	0.042	-0.023	0.081	0.099	0.044	0.12	0.018	0.036
policy_annual_premium	0.061	0.0016	-0.043	-0.074	1	0.0064	0.035	-0.034	0.046	-0.035	-0.035	0.016	-0.045	-0.11	-0.027	-0.072	-0.037	-0.083
umbrella_limit -	0.015	0.029	0.014	0.0057	0.0064	1	0.015	-0.044	-0.031	-0.034	-0.017	0.01	-0.017	-0.016	-0.051	-0.042	-0.022	0.00038
insured_zip -	0.039	0.026	-0.02	-0.041	0.035	0.015	1	0.0072	-0.0048	0.0073	0.008	0.025	0.032	-0.062	0.0039	-0.012	-0.026	-0.052
capital-gains -	-0.0022	-0.014	0.02	0.029	-0.034	-0.044	0.0072	1	-0.021	-0.013	0.058	0.074	-0.026	0.068	0.022	0.016	0.061	0.042
capital-loss -	0.027	-0.003	-0.013	0.0032	0.046	-0.031	-0.0048	-0.021	1	-0.023	-0.0014	-0.025	-0.036	0.007	-0.047	0.017	-0.017	-0.044
incident_hour_of_the_day -	0.057	0.08	0.088	0.072	-0.035	-0.034	0.0073	-0.013	-0.023	1	0.12	-0.034	0.00055	0.16	0.16	0.13	0.2	0.019
number_of_vehicles_involved -	0.0054	0.014	0.034	0.042	-0.035	-0.017	0.008	0.058	-0.0014	0.12	1	0.0018	-0.015	0.17	0.21	0.18	0.22	0.038
bodily_injuries	-0.0015	-0.0011	0.03	-0.023	0.016	0.01	0.025	0.074	-0.025	-0.034	0.0018	1	0.0041	0.066	0.055	0.028	0.051	-0.019
witnesses -	0.045	0.07	-0.0075	0.081	-0.045	-0.017	0.032	-0.026	-0.036	0.00055	-0.015	0.0041	1	0.036	-0.011	0.11	-0.02	0.062
total_claim_amount -	0.008	0.077	0.006	0.099	-0.11	-0.016	-0.062	0.068	0.007	0.16	0.17	0.066	0.036	1	0.53	0.57	0.88	0.026
injury_claim ·	0.058	0.067	-0.0019	0.044	-0.027	-0.051	0.0039	0.022	-0.047	0.16	0.21	0.055	-0.011	0.53	1	0.44	0.61	-0.012
property_claim -	-0.0096	0.073	0.0097	0.12	-0.072	-0.042	-0.012	0.016	0.017	0.13	0.18	0.028	0.11	0.57	0.44	1	0.5	0.046
vehicle_claim -	0.038	0.055	0.049	0.018	-0.037	-0.022	-0.026	0.061	-0.017	0.2	0.22	0.051	-0.02	0.88	0.61	0.5	1	-0.014
auto_year ·	-0.011	0.0071	0.041	0.036	-0.083	0.00038	-0.052	0.042	-0.044	0.019	0.038	-0.019	0.062	0.026	-0.012	0.046	-0.014	1
_c39 ·																		
	months_as_customer -	- aĝe	policy_number -	policy_deductable -	policy_annual_premium -	umbrella_limit -	insured_zip -	capital-gains -	capital-loss -	ncident_hour_of_the_day -	ber_of_vehicles_involved -	bodily_injuries -	witnesses -	total_claim_amount -	injury_claim -	property_claim -	vehicle_claim -	auto_year -





## Results

By using Gradient
 Boosting Algorithm, we obtain a accuracy of 83.33%



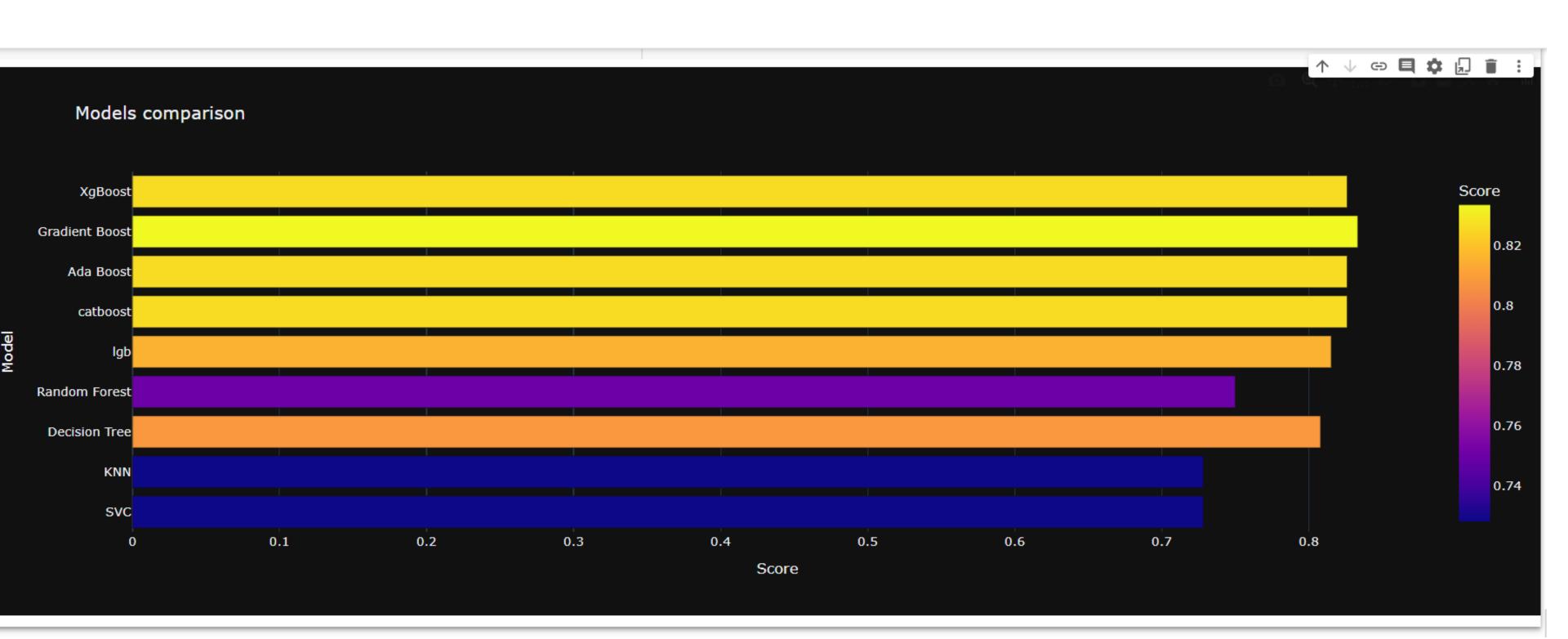
₽





7Gradient Boost0.8333335catboost0.8260876Ada Boost0.8260878XgBoost0.8260874lgb0.8152172Decision Tree0.8079713Random Forest0.7500000SVC0.7282611KNN0.728261			
<ul> <li>Ada Boost 0.826087</li> <li>XgBoost 0.826087</li> <li>Igb 0.815217</li> <li>Decision Tree 0.807971</li> <li>Random Forest 0.750000</li> <li>SVC 0.728261</li> </ul>	7	Gradient Boost	0.833333
<ul> <li>XgBoost 0.826087</li> <li>Igb 0.815217</li> <li>Decision Tree 0.807971</li> <li>Random Forest 0.750000</li> <li>SVC 0.728261</li> </ul>	5	catboost	0.826087
<ul> <li>4 Igb 0.815217</li> <li>2 Decision Tree 0.807971</li> <li>3 Random Forest 0.750000</li> <li>O SVC 0.728261</li> </ul>	6	Ada Boost	0.826087
<ul> <li>Decision Tree 0.807971</li> <li>Random Forest 0.750000</li> <li>SVC 0.728261</li> </ul>	8	XgBoost	0.826087
3 Random Forest 0.750000	4	lgb	0.815217
<b>o</b> SVC 0.728261	2	Decision Tree	0.807971
	3	Random Forest	0.750000
1 KNN 0.728261	0	SVC	0.728261
	1	KNN	0.728261

# Model Comparison



# Thank You