Car Insurance & Machine Learning

Pieter Slabber

Table of content

1. The steak holder and the problem.
2. Introduction to the data
3. Visualizations
4. Machine Learning Model
5. Model Evaluation
6. Recommendation

1. Steak holder and the problem

Context:

The insurance company has shared its annual car insurance data. Now, I have to find out the real customer behaviors over the data.

Content:

The columns are resembling practical world features.

The ‘OUTCOME’ column indicates 1 if a customer has claimed his/her loan else 0.

Problem or challenge:

The steak holder needs to know whether a client will claim or not.

This type of problem in machine learning is called a binary classification problem.

2. Introduction to the data

Summary of the data:

The dataset consists of 19 features from there 18 of them are corresponding logs which were taken by the insurance company. 10000 records were used in the dataset

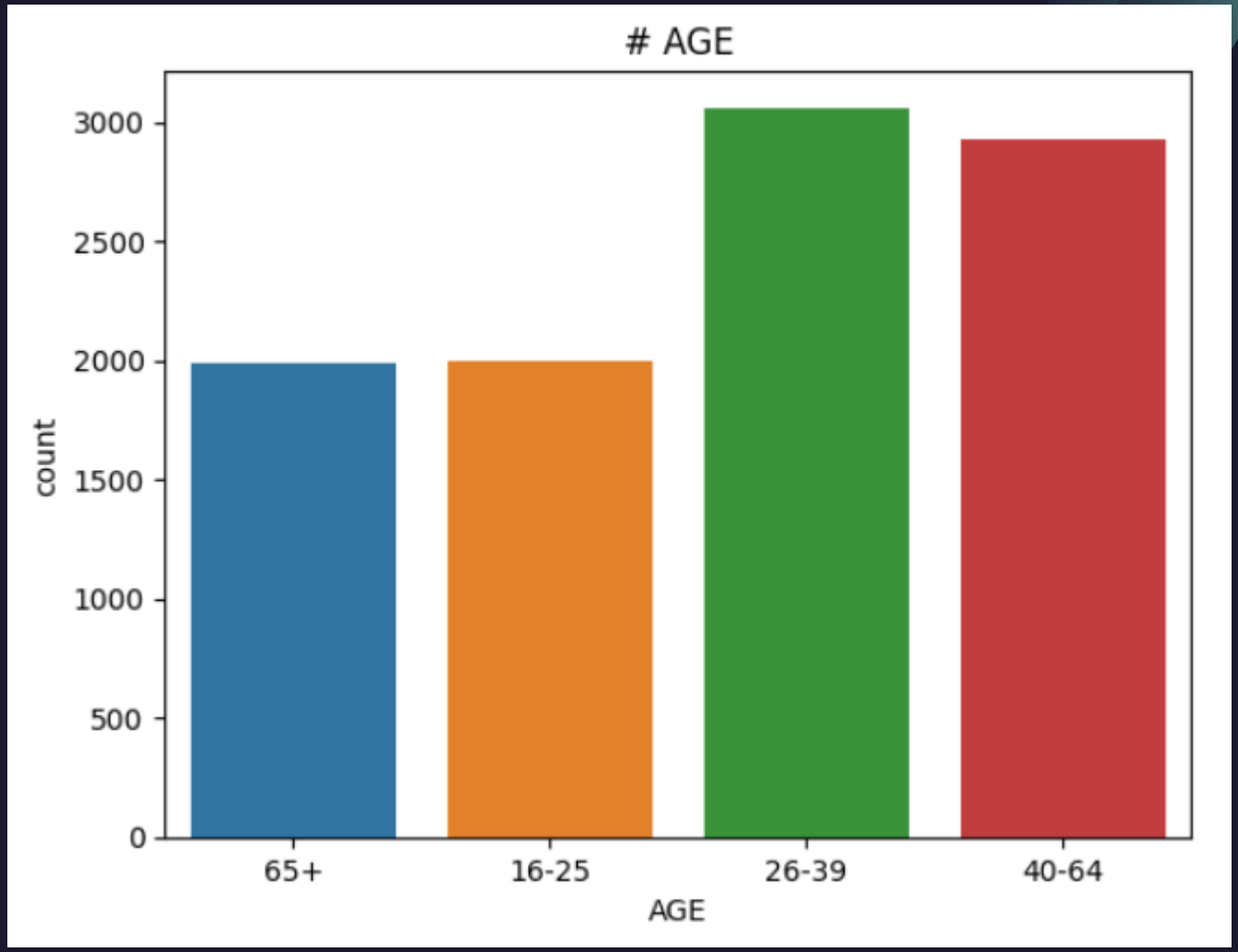
The target is simple terms is the output column the machine learning needs to predict. In this case it is the 'Outcome' column.

Data columns: AGE, GENDER, RACE, EDUCATION, INCOME, CREDIT_SCORE, VEHICLE_OWNERSHIP, VEHICLE_YEAR, MARRIED, CHILDREN, POSTAL_CODE, ANNUAL_MILEAGE, VEHICLE_TYPE, SPEEDING_VIOLATIONS, DUIS, PAST_ACCIDENTS, OUTCOME

AGE	GENDER	RACE	DRIVING	EDUCATION	INCOME	CREDIT_SCORE	VEHICLE_OWNERSHIP	VEHICLE_YEAR	MARRIED	CHILDREN	POSTAL_CODE	ANNUAL_MILEAGE	VEHICLE_TYPE	SPEEDING_VIOLATIONS	DUIS	PAST_ACCIDENTS	OUTCOME
16-25	female	majority	0-9y	high school	poverty	0.629027314	1	before 2015	0	0	10238	12000	sedan	0	0	0	0
26-39	male	minority	10-19y	none	working class	0.357757117	0	after 2015	1	1	32765	16000	sports car	1	1	1	1
40-64			20-29y	university	middle class	0.493145785					21217	11000		2	2	2	
65+			30y+		upper class	0.206012851						11000		3	3	3	

3. Visualizations

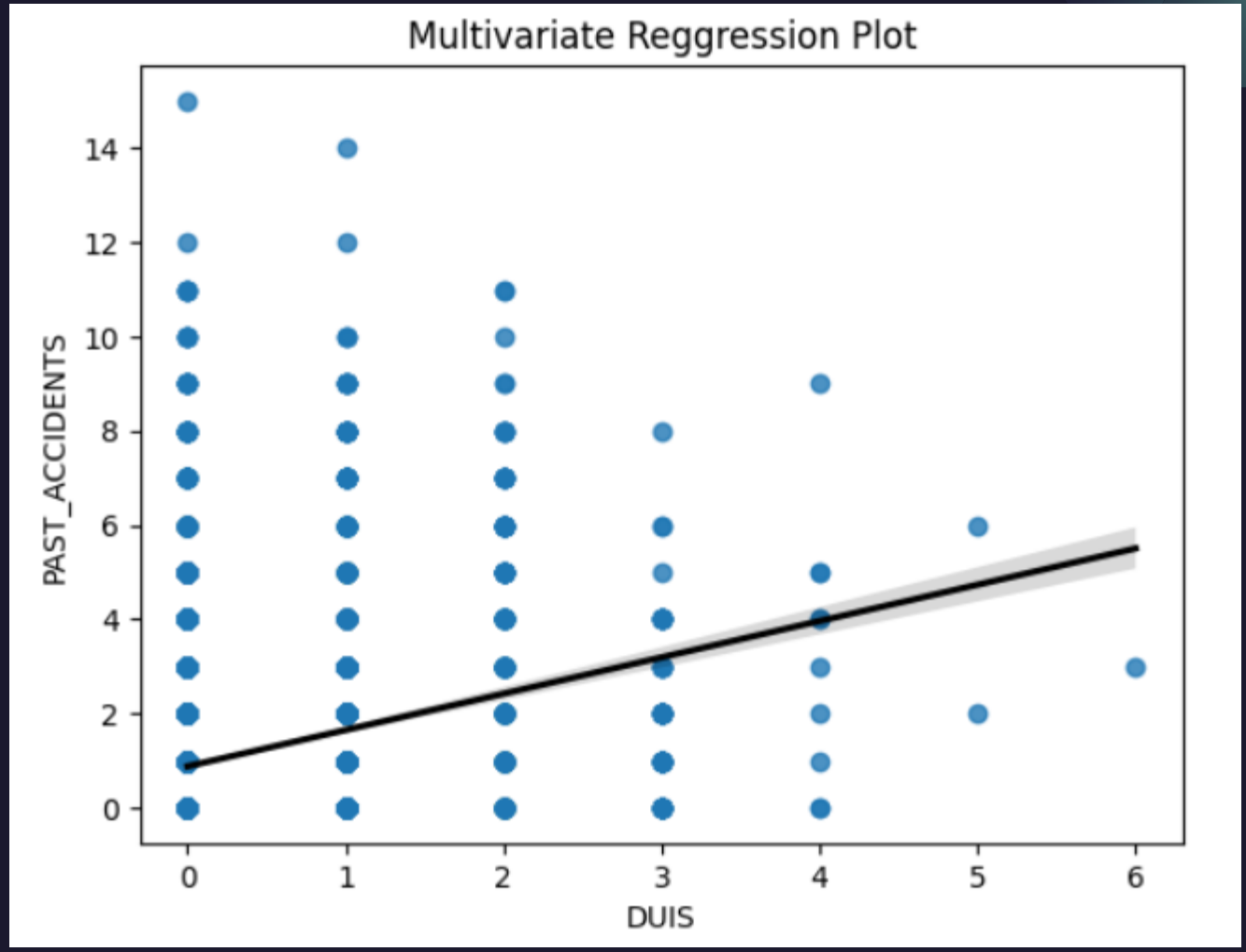
This visualization displays the age group distribution in the dataset. The 65+ and the 16 – 25 groups are the same amount of people. The 26 - 30 and 40 – 64 groups are the highest age groups represented.



Visualizations - continue

This visualization displays the policy holders past accidents relationship with the policy holders who were driving under the influence.

One can see clearly that the accidents count gets higher as the duis increases.



4. Machine learning model

The machine learning model used for the prediction is called a KNN model.

This model is normally easy to implement.

Performance may degrade as the number of features increase.

You may need more computing power on large datasets.

5. Model evaluation

The model predicted a lower number of false positives than false negatives.

This means that the number of policy holders predicted to claim and will not claim are lower than the ones predicted not to claim and eventually will claim.

This could be problem because there might be more policy holders that will claim than what was predicted.

6. Recommendation

Machine learning is transforming the car insurance industry by making it more efficient.

- By analyzing data on things like driving history and age, machine learning algorithms can create models that suggest policies with added features or benefits that may interest customers.
- Machine learning can also be used to identify high-risk drivers in the car insurance industry. By analyzing data on things like driving history and accidents, machine learning algorithms can create models that identify drivers who are more likely to have an accident. This information can be used by insurers to refuse cover or increase premiums for high-risk drivers
- As a result of these efforts, car insurance should become more affordable and accessible for consumers worldwide.

Thank You